







```

FH Key      Location/Qualifiers
FT Peptide  1..19
FT Protein  /label= Signal_peptide
FT          20..431
XX          /note= "Mature pNOV4054 phytase fusion protein"
FN WO2003057248-A1.
PD 17-JUL-2003.
XX
XX 30-DEC-2002; 2002WO-US0411787.
XX
XX 28-DEC-2001; 2001US-0344476P.
XX
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
XX Lanahan MB, Betts S;
XX
XX WPI; 2003-607980/57.
XX N-PSDB; AAD57148.
XX
XX Preparing a thermotolerant phytase for preparing animal feed or human
XX food by expressing in a plant cell an expression cassette comprising a
XX promoter operably linked to a nucleic acid molecule encoding a
XX thermotolerant phytase.
XX
XX Claim 63; Page 90; 157pp; English.
XX
XX The invention relates to a method for preparing a thermotolerant phytase.
XX The method comprises expressing in a plant cell an expression cassette
XX comprising a promoter operably linked to a nucleic acid molecule encoding
XX a thermotolerant phytase which retains at least 40% activity after 30
XX minutes at 60plusC and has a specific activity of greater than 200 U/mg
XX at pH 4.5 and 37plusC. The method is useful for preparing a
XX thermotolerant phytase for preparing animal feed or human food. The
XX invention is useful for reducing the feed conversion ratio and increasing
XX weight gain of animals fed diets with inorganic phosphate at levels below
XX 0.45%, minimising dietary requirements of phosphorus in an animal,
XX enhancing the utilisation of phosphorus present in animal feed, enhancing
XX organic phosphorus utilisation from organic phosphorus sources in animal
XX feed, decreasing the phosphate levels in excreta from an animal,
XX improving the processing of grain, improving the nutritive value of
XX processed grain product or a method of processing grain, improving the
XX nutritive value of animal feed and human food, and preparing a
XX transformed plant which expresses a thermotolerant phytase. The present
XX sequence is pNOV4054 phytase fusion protein used in the exemplification
XX of the invention
XX
XX Sequence 431 AA;
XX
XX Query Match      39, 8%; Score 655.5; DB 6; Length 431;
XX Best Local Similarity 46.2%; Pred. No. 6.2e-56;
XX Matches 198; Conservative 43; Mismatches 75; Indels 113; Gaps 34;
XX
XX QY 1 VSLALLAALLAALAPAAAAP-LKLEKVVILSRHGRVSPTKA--LM--VSPD-WP-WP 53
XX   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX Db 3 VLLVALALLAAGATSAAQSEPELKLESVVIVSRHGVRAPTKATQLMQVDTPDAWTPW 62
XX
XX QY 54 V---GLTPRGAALV-LIGY-----RYFARGILP--CCPAAG--TILADVDERTR-TGQAF 100
XX   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX Db 63 VKLGELTRGELIAYLGHYWRQLVADGLLPKCGCPOSGOVALIADVDERTRTKTGEAFA 122
XX
XX QY 101 AGLAPGCAI-VH-----AGDDPIIFHGLDTG-C-LDQA---DAILAAGEGGITARHG---- 145
XX   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX Db 123 AGLAPDCAITVHTQADTSSPDLEPLNPLKTVGVCQLDNANVTDAILLERAGGSIADPTGHYQT 182
XX
XX QY 146 --LTIKVLNF--ASACI-----ECARVVGRIGLPILLASTLSE 178
XX   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX Db 183 AFRELERVLNFPQNLCKREKQDESCSLTQALPSELKVGADCVSLTGAVS--LASMLTE 240
XX
XX QY 179 IFLELYAQGPMEVWGRI--SAEW-----LLHNAQF-LMNRTPYIA--RAYPI--LIVTAL 227
XX   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

---

```

Db 241 IFLLQAQGMPEPGWGRITDSHQWNTLLSLHNAQFDLLQRTPEVARSRATPLLDLTKTAL 300
QY 228 S-----PARVVLLAGHTNLA-LGG-LDL-WQLP-QDETTPGALVPELWN 270
Db 301 TPHPQKQAYGVTLPSTVLFIAGHDTNLNLGCALELANWTLFGQPDNTPPGGELYFERWR 360
QY 271 R-----YRVMTYQTMQDLNLEPLLPRIPL-----IPGC---GSEAAACSLSDPARL 313
Db 361 RLSDNSQMTQVSLVFQTLQOMEDKTFLSINTPPGGEVKLTLAGCEERNAQKCSLAGFTQI 420
QY 314 V-----APAC 318
Db 421 VNEARIPAC 429
XX
XX RESULT 4
XX ID AAE37854 standard; protein; 437 AA.
XX AC AAE37854;
XX
XX DT 06-NOV-2003 (first entry)
XX DE pNOV4058 phytase fusion protein.
XX
XX KW Thermotolerant phytase; weight gain; animals fed diet; grain processing;
XX nutritive value; transformed plant; anabolic; chimeric; maize.
XX
XX OS Chimeric - Zea mays.
XX OS Chimeric - Unidentified.
XX
XX FH Key      Location/Qualifiers
XX FT Peptide  1..19
XX FT Protein  /label= Signal_peptide
XX FT          20..437
XX FT          /note= "Mature pNOV4058 phytase fusion protein"
XX FT Region   432..437
XX FT          /note= "SEKDEL signal"
XX
XX WO2003057248-A1.
XX
XX 17-JUL-2003.
XX
XX 30-DEC-2002; 2002WO-US0411787.
XX
XX 28-DEC-2001; 2001US-0344476P.
XX
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
XX Lanahan MB, Betts S;
XX
XX WPI; 2003-607980/57.
XX N-PSDB; AAD57149.
XX
XX Preparing a thermotolerant phytase for preparing animal feed or human
XX food by expressing in a plant cell an expression cassette comprising a
XX promoter operably linked to a nucleic acid molecule encoding a
XX thermotolerant phytase.
XX
XX Example 1; Page 91; 157pp; English.
XX
XX The invention relates to a method for preparing a thermotolerant phytase.
XX The method comprises expressing in a plant cell an expression cassette
XX comprising a promoter operably linked to a nucleic acid molecule encoding
XX a thermotolerant phytase which retains at least 40% activity after 30
XX minutes at 60plusC and has a specific activity of greater than 200 U/mg
XX at pH 4.5 and 37plusC. The method is useful for preparing a
XX thermotolerant phytase for preparing animal feed or human food. The
XX invention is useful for reducing the feed conversion ratio and increasing
XX weight gain, improving reducing feed conversion ratios or increasing
XX weight gain of animals fed diets with inorganic phosphate at levels below
XX 0.45%, minimising dietary requirements of phosphorus in an animal,
XX enhancing the utilisation of phosphorus present in animal feed, enhancing

```



```

XX DE Escherichia coli appA phytase mutant protein.
XX KW Bacterial phytase; K12 appA phytase; protease stability; anabolic;
XX KW gastrointestinal; nutritional value; feed treatment process; therapy;
XX KW thermal tolerance; growth performance; alcoholic drink; biopulping;
XX KW non-alcoholic drink; biobleaching; mutant; muten.
XX OS Escherichia coli.
XX OS Synthetic.
XX FH Key Location/Qualifiers
FT Misc-difference 68 /note= "Wild type Trp substituted with Glu"
FT Misc-difference 84 /note= "Wild type Gln substituted with Trp"
FT Misc-difference 95 /note= "Wild type Ala substituted with Pro"
FT Misc-difference 97 /note= "Wild type Lys substituted with Cys"
FT Misc-difference 168 /note= "Wild type Ser substituted with Glu"
FT Misc-difference 180 /note= "Wild type Arg substituted with Tyr"
FT Misc-difference 225 /note= "Wild type Asn substituted with Cys"
FT Misc-difference 276 /note= "Wild type Tyr substituted with Asp"
XX WO200190333-A2.
XX 29-NOV-2001.
XX 24-MAY-2001; 2001WO-US017118.
XX 25-MAY-2000; 2000US-00580515.
XX (DIVE-) DIVERSA CORP.
XX Short JM, Kretz KA, Gray KA, Barton NR, Garrett JB, O'donoghue E;
XX WPI; 2002-083108/11.
XX New bacterial phytase for e.g. improving the nutritional value of phytate
XX -containing foodstuffs and subsequently improving the growth performance
XX of an organism that consumes it, or in treating animal digestive systems.
XX Claim 59; Fig 8; 170pp; English.
XX The patent discloses recombinant bacterial phytase from Escherichia coli
XX K12 appA phytase. The enzyme has phytase activity and improved thermal
XX tolerance when compared with wild-type phytase. It has improved protease
XX stability at low pH. The recombinant phytase is useful for improving the
XX nutritional value of phytate-containing foodstuffs and subsequently
XX improving the growth performance of an organism that consumes it, in
XX treating animal digestive systems, in feed treatment processes and for in
XX vitro purposes related to research, discovery and development. They are
XX also used for generating recombinant digestive system life forms, for
XX producing or manufacturing alcoholic and non-alcoholic drinks based on
XX the use of moulds, grains and/or plants, in biopulping and bio-bleaching
XX where a reduction in the use of environmentally harmful chemicals that
XX are traditionally used in the pulp and paper industry is desired and in
XX the reduction or possible elimination of the need for mineral
XX supplements, enzymes or therapeutic drugs for animals from the daily feed
XX thus increasing the amount calories and nutrients present in the feed.
XX The present sequence is E. coli appA phytase mutant protein
SQ Sequence 430 AA;
Query Match 38.4%; Score 632.5; DB 5; Length 430;
Best Local Similarity 45.2%; Pred. No. 1.2e-53;
Matches 194; Conservative 44; Mismatches 76; Indels 115; Gaps 35;

```

```

QY 2 SLLLAALLAALAP--AAAAEP-LKLEKVVILSRHVSPTKA--LM--VSPD-WP-W 52
DB 3 AILPFLSLIPLTPQSAFAQSEPELKLESVIVSRHGVRAPTKATQLMQDVTFDANPTW 62
QY 53 PV---GLTPRGAALV-LIGY---RYFARGLLP--GCPAAG--TILADVDERTR-TGOAF 99
DB 63 PVKLGELTPRGELIAYLGHYWRQELVADGILLPCGCPQSGQVAILADVDERTRKTGEAF 122
QY 100 AAGLAPGCAI-VH-----AGDDPIFHGLDTG-C-LDQA---DAILAAGEGLTARHG--- 145
DB 123 AAGLAPDCAITVHTQADTSSPDPLFNPLKTVGCQLDNANVTDAILEAG-GSIADPTGRYQ 181
QY 146 ---LTLAKLVNF--ASACL-----ECARVVVGRIGPLIASTLS 177
DB 182 TAFRELRLVNFPPQSNLCKREKQDECSLTQALPSELKVSADCVSLTGAVS--LASMLT 239
QY 178 EIFLLEYAQGMEVGVNGRI--SAEW-----LHNAQFLNRTPIYA--RATPI--LIVTAL 227
DB 240 EIFLLQQAQGMPEFGWGRITDSHOWNTLLSLHNAQFDLQRTPEVARSATPLLDIKTAL 299
QY 228 S-----PARVVLGHADTNLA-LGG-LDL-WQLP-OPDETPPGCALVFFELWN 270
DB 300 TPHPQKQAYGVTLTSTVPIAGHDTNLNANLGGALNLNWTLPQGDDNTPPGGELYFERWR 359
QY 271 R-----YVRVMYQTMQDLRNLEPLRLP-----IPGC---GSEAAACSLSDPARL 313
DB 360 RLSDNSQWIVSLVFTLQQRDKTPLSLNTPPGVEVKLTLAGCBERNAGMCSLAGTQI 419
QY 314 V---APAC 318
DB 420 VNEARIPAC 428
RESULT 7
ADA19451
ID ADA19451 standard; protein; 430 AA.
XX ADA19451;
XX 20-NOV-2003 (first entry)
DE E. coli K12 phytase mutant.
XX Phytase; enzyme; phytate; appA gene; animal feed; inorganic phosphate;
XX digestion enhancement; transgenic; thermal tolerance; protease stability;
XX mutant; muten.
XX Synthetic.
XX Escherichia coli; strain K12.
XX Key Location/Qualifiers
FH Misc-difference 68 /note= "Wild-type Trp substituted by Glu"
FT Misc-difference 72 /note= "Encoded by CGN"
FT Misc-difference 84 /note= "Wild-type Gln substituted by Trp"
FT Misc-difference 95 /note= "Wild-type Ala substituted by Pro"
FT Misc-difference 97 /note= "Wild-type Lys substituted by Cys"
FT Misc-difference 168 /note= "Wild-type Ser substituted by Glu"
FT Misc-difference 180 /note= "Wild-type Arg substituted by Tyr"
FT Misc-difference 226 /note= "Wild-type Asn substituted by Cys"
FT Misc-difference 277 /note= "Wild-type Tyr substituted by Asp"
XX US2002136754-A1.
XX 26-SEP-2002.

```

```

XX PF 24-MAY-2001; 2001US-00866379.
XX PR 13-AUG-1997; 97US-00910798.
XX PR 01-MAR-1999; 99US-00259214.
XX PR 13-APR-1999; 99US-00291931.
XX PR 25-MAY-1999; 99US-00318528.
XX PR 25-MAY-2000; 2000US-00580515.
XX (SHOR/) SHORT J M.
XX PA (KRET/) KRETZ K A.
XX PA (GRAY/) GRAY K A.
XX PA (BART/) BARTON N R.
XX PA (GARRE/) GARRETT J B.
XX PA (DONO/) O' DONOGHUE E.
XX PA (MATH/) MATHUR E J.
XX PI Short JM, Kretz KA, Gray KA, Barton NR, Garrett JB;
XX PI O' Donoghue E, Mathur EJ;
XX XX WPI; 2003-040002/03.
XX XX
XX XX Isolated Escherichia coli polynucleotide encoding a modified phytase
XX PT enzyme, useful in the production of animal feed, for improving the
XX PT nutritional value of phytate-containing feedstuff and for enhancing
XX PT digestion in humans and animals.
XX XX
XX PS Claim 59; Fig 8; 62pp; English.
XX XX
XX CC The invention relates to an isolated Escherichia coli polynucleotide
XX CC encoding a phytase enzyme appearing as ADA19450 and having amino acids
XX CC modification from W68E, Q84W, A95P, K97C, S168E, R180Y, N226C or Y277D.
XX CC Also included the E. coli appA gene ADA19449 (or an oligonucleotide
XX CC derived from it) or its mutant sequence ADA19452, expression vectors,
XX CC host cells, a method of improving nutritional value of a phytate-
XX CC containing feedstuff by contacting the phytate-containing feedstuff with
XX CC a pure phytase enzyme mutated phytase (where the phytase enzyme catalyzes
XX CC the liberation of inorganic phosphate from the phytate in the phytate-
XX CC containing feedstuff), a method to produce an animal feed containing a
XX CC microbial phytase (comprising culturing the plant cell, plant part or
XX CC plant under conditions where the nucleotide sequence is expressed and
XX CC converting the plant cells, plants or plant into a composition for animal
XX CC feed), a feed composition for animals (comprising the plant seeds, plant
XX CC cells, plant parts or plants in admixture with a phytate-containing
XX CC feedstuff), a method to treat a human or an animal able to benefit from
XX CC digestive enhancement by the activity of an exogenous phytase enzymes
XX CC comprising administering to the human or animal the plant seed, plant
XX CC cells, plant parts or plants of a transgenic plant which is modified to
XX CC contain an expression system which expresses a nucleotide sequence
XX CC encoding a phytase enzyme, a transgenic non-human organism whose genome
XX CC comprising a heterologous nucleic acid sequence encoding a polypeptide
XX CC having phytase activity. The phytase enzyme is useful for improving the
XX CC nutritional value of phytate-containing feedstuff, in the production of
XX CC animal feed and for enhancing digestion in humans and animals. The
XX CC invented method improves thermal tolerance and protease stability. It
XX CC also improves the feeding value of phytate rich ingredients. The present
XX CC sequence represents E. coli K12 mutant phytase.
XX XX
XX XX Sequence 430 AA;
XX XX
XX XX Query Match 38.4%; Score 632.5; DB 6; Length 430;
XX XX Best Local Similarity 45.2%; Pred. No. 1.2e-53;
XX XX Matches 194; Conservative 44; Mismatches 76; Indels 115; Gaps 35;
XX QY 2 SLALLAALAAAP-AAAAAP-LKLEKVTLSRHGVSPKA--LM--VSPD-WP-W 52
XX Db 3 AILIPFLSLIPTPQAFQAFQAFQAFQAFQAFQAFQAFQAFQAFQAFQAFQAF 62
XX QY 53 PV---GLTPRGALV-LLGY---RVFARGLLP--GCPAAG--TILADVDERTR-TGQAF 99
XX Db 63 PVKLGLTPRGGLIAYLGHYWRQRLVADGLLPCGPGQGVALLIADVDERTRKTGEAF 122
XX QY 100 AAGLAPCAITVHTQADTSSPDLFNPDKTGVQCLDNANTVDALEAG-GSIADFTGHYQ 181

```

```

Db 123 AAGLAPCAITVHTQADTSSPDLFNPDKTGVQCLDNANTVDALEAG-GSIADFTGHYQ 181
QY 146 ---LTLAKVLNF--ASACL-----ECARVVGSLGFLLASTLS 177
Db 182 TAFEELERVLNFPQSNLCKREKQDSCLTQALPSGLKVSADCVSLTGAVS--LASMLT 239
QY 178 EIFLLEVAQCPMEVGRI--SAEM-----LLHNAQPLMRTPIYA--RATPI--LIVTAL 227
Db 240 EIFLLQAQGNPFGWGRITDSHQWNTLLSLHNAQFDLQRTPEVARSRATPLDLIKTAL 299
QY 228 S-----PARVILLAGHDNTLA-LGG-LDL-WOLP-QPDETTPGGALVFELMN 270
Db 300 TPHPPOKAYGVTLPTSLVFIAGHDNTLANLNGALELNTLPGQPDNTPPGSELVFERWR 359
QY 271 R-----YRVVMYQTMQDLNLEPLPRILP-----IPGC---GSEACSLSDPARL 313
Db 360 RLSDNSQWIQVSLVFQTLQQRDKTPLUSLNTPPGEVKLTLAGCERNNAQGMCSLAGFTQI 419
QY 314 V-----APAC 318
Db 420 VNEARIIPAC 428
XX
XX RESULT 8
XX ID ADC87743 standard; protein; 436 AA.
XX AC ADC87743;
XX DT 01-JAN-2004 (first entry)
XX DE Modified Escherichia coli phytase.
XX KW Phytase; food supplement; enzyme delivery matrix; soybean meal;
XX KW thermotolerance; thermostability; kernel; phytate;
XX KW myo-inositol-hexaphosphate; inositol; inorganic phosphate;
XX KW thermotolerant; feed value; digestion; enzyme.
XX OS Escherichia coli.
XX OS Escherichia coli.
XX PN US2003103958-A1.
XX PD 05-JUN-2003.
XX XX 24-MAY-2002; 2002US-00156660.
XX PR 13-AUG-1997; 97US-00910798.
XX PR 01-MAR-1999; 99US-00259214.
XX PR 13-APR-1999; 99US-00291931.
XX PR 25-MAY-1999; 99US-00318528.
XX PR 25-MAY-2000; 2000US-00580515.
XX PR 24-MAY-2001; 2001US-00866379.
XX PA (DIVE-) DIVERSA CORP.
XX PI Short JM, Kretz K, Gray KA, Barton NR, Garrett JB, O' donoghue E;
XX PI Mathur EJ;
XX XX WPI; 2003-787039/74.
XX DR N-PSDB; ADC87742.
XX XX
XX PT New nucleic acid encoding a polypeptide having phytase activity, useful
XX PT in improving the feeding value of phytate rich ingredients or as an aid
XX PT in phytate digestion.
XX XX
XX PS Claim 60; SEQ ID NO 2; 113pp; English.
XX XX
XX CC The invention discloses a new isolated or recombinant nucleic acid which
XX CC encodes a polypeptide having a phytase activity. Also claimed is a
XX CC nucleic acid probe, an amplification primer sequence pair, an expression
XX CC cassette comprising the nucleic acid, a vector comprising the nucleic

```

acid, a transgenic non-human animal or plant, or its seed, comprising the nucleic acid, an antisense oligonucleotide, inhibiting the translation of a phytase message in a cell, a heterodimer comprising the polypeptide and a second domain, an array comprising immobilised polypeptide or nucleic acid, a hybridoma comprising an antibody that specifically binds to the polypeptide, a food supplement for an animal, an edible enzyme delivery matrix, an edible pellet comprising a granule edible carrier and the polypeptide, a feed composition, a soybean meal, isolating or identifying the polypeptide, making an anti-phytase antibody, producing a recombinant polypeptide, determining whether a compound binds to the polypeptide, identifying a modulator, whole cell engineering of new or modified phenotypes by using real-time metabolic flux analysis, increasing thermotolerance or thermostability of the phytase polypeptide, increasing the resistance of the phytase polypeptide to enzymatic inactivation in a digestive system and processing of corn and sorghum kernels. The phytase activity comprises catalysis of phytate (myo-inositol-hexaphosphate) to inositol and inorganic phosphate and is thermotolerant. The nucleic acid is useful in improving the feeding value of phytate rich ingredients or as an aid in phytate digestion. The sequence presented is the modified *Escherichia coli* phytase.

Query Match 39.1%; Score 627.5; DB 7; Length 436;  
 Best Local Similarity 44.9%; Pred. No. 3.7e-53;  
 Matches 193; Conservative 44; Mismatches 78; Indels 115; Gaps 35;

Sequence 436 AA;

2 SLLAALLAALAP--AAAAEP-LKLEKVVILSRHGVSPKA--LM--VSPD-WP-W 52  
 3 ALLPFLSLIPLTPQSAFAQSEPELKLESVIVSRHGVRAPTKATQMDVTPDAWPTW 62  
 53 PV---GLTPRGAALV-LIGY----RYFARGLLP--GCPAAG--TILADVDETR--TGOAF 99  
 63 PVKLGELTPRGGLIAYLGHYWRQLRVADGLLPKCGCPQSGQVAIIADVDETRTKTGAF 122  
 100 AAGLAPGCAI-VH-----AGDDPIFHGLDTG-C-LDOA-----DAILAAGEGLTARHG--- 145  
 123 AAGLAPDCAITVHTQADTSPPDPLFNKGTGVCOLDNANVTDAILERAGGSIADFTGHYQ 182  
 146 ---LTLAKVLF--ASACL-----ECLAVVGRGLTARHG---ASACL 159  
 183 TAFELSERVNFPSNLCIKREKODESCSLTQALPSBLKVSADCVSLTGAVS--LASMLT 240  
 178 EIFLLEYAQGMPEWGRI--SAEW----LLHNAQF--LMRTPIA--RATPI--LIVTALS 232  
 241 EIFLLOQAQGMPEWGRI--TDSHWNLTLLSHNAQFDLLQRTPEVARSATPLDILIKTA 300  
 227 LS-----PARVLLAGHDNTLA-LGG-LDL-WOLP--QPDETPPGGALVPELW 269  
 301 LTPHPKQAYGVTLPTSLVFLTGHAGDNTLANLGALELNTLFGQPDPTPPGGELVFERW 360  
 270 NR-----YVRVYQTMQRLNLEPLPRILP-----IPGC-----GSEAAACSLGDFAR 312  
 361 RLSDNSQWTCVSLVFTLQOMRKPTPLSLNTPPGEVKVLTLAGEERNAQGMCSLAGFTQ 420  
 313 LV-----APAC 318  
 421 IVNEARIPAC 430

# RESULT 9

AAE37851  
 ID AAE37851 standard; protein; 412 AA.  
 XX  
 AC AAE37851;  
 DT 06-NOV-2003 (first entry)  
 XX  
 DE Maize-optimised Nov9X phytase mutant.  
 XX  
 KW Thermotolerant phytase; weight gain; animals fed diet; grain processing;  
 XX nutritive value; transformed plant; anabolic; maize; mutant; muten.

OS Zea mays.  
 OS Synthetic.  
 XX  
 PN WO2003057248-A1.  
 XX  
 PD 17-JUL-2003.  
 XX  
 PF 30-DEC-2002; 2002WO-US041787.  
 XX  
 PR 28-DEC-2001; 2001US-0344476P.  
 XX  
 PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
 XX  
 PI Lanahan MB, Betts S;  
 XX  
 DR WPI; 2003-607980/57.  
 XX  
 DR N-PSDB; AAD57147.  
 XX  
 PT Preparing a thermotolerant phytase for preparing animal feed or human food by expressing in a plant cell an expression cassette comprising a promoter operably linked to a nucleic acid molecule encoding a thermotolerant phytase.  
 PT  
 PT Claim 3; Page 89; 157pp; English.  
 PS  
 XX  
 CC The invention relates to a method for preparing a thermotolerant phytase. The method comprises expressing in a plant cell an expression cassette comprising a promoter operably linked to a nucleic acid molecule encoding a thermotolerant phytase which retains at least 40% activity after 30 minutes at 60pluoc and has a specific activity of greater than 200 U/mg at pH 4.5 and 37pluoc. The method is useful for preparing a thermotolerant phytase for preparing animal feed or human food. The invention is useful for reducing the feed conversion ratio and increasing weight gain, improving reducing feed conversion ratios or increasing weight gain of animals fed diets with inorganic phosphate at levels below 0.4%, minimising dietary requirements of phosphorus in an animal, enhancing the utilisation of phosphorus present in animal feed, enhancing organic phosphorus utilisation from organic phosphorus sources in animal feed, decreasing the phosphate levels in excreta from an animal, improving the processing of grain, improving the nutritive value of processed grain product or a method of processing grain, improving the nutritive value of animal feed and human food, and preparing a transformed plant which expresses a thermotolerant phytase. The present sequence is maize-optimised Nov9X phytase mutant used in the exemplification of the invention

Query Match 37.4%; Score 616.5; DB 6; Length 412;  
 Best Local Similarity 45.5%; Pred. No. 4.2e-52;  
 Matches 187; Conservative 41; Mismatches 70; Indels 113; Gaps 34;

19 AAAEP-LKLEKVVILSRHGVSPKA--LM--VSPD-WP-WPV---GLTPRGAALV-LIG 67  
 2 AQSEPELKLESVIVSRHGVRAPTKATQMDVTPDAWPTVFKLGELTPRGGLIAYIG 61  
 68 Y----RYFARGLLP--GCPAAG--TILADVDETR--TGOAFAAGLAPGCAI-VH-----A 112  
 62 HYWRQLRVADGLLPKCGCPQSGQVAIIADVDETRTKTGAEFAAGLAPDCAITVHTQADTS 121  
 113 GDDPIFHGLDTG-C-LDOA-----DAILAAGEGLTARHG-----LTLAKVLF--ASACL 159  
 122 SPDDLNFNPLKGTGVCOLDNANVTDAILERAGGSIADFTGHYQTAFAELSERVNFPSNLC 181  
 160 -----ECARVVGRGLTARHG-----LTLAKVLF--ASACL 159  
 182 KREKODESCSLTQALPSBLKVSADCVSLTGAVS--LASMLTEIFLLOQAQGMPEWGRI 239  
 197 --SAEW----LLHNAQF--LMRTPIA--RATPI--LIVTALS-----PARV 232  
 240 TDSHWNLTLLSHNAQFDLLQRTPEVARSATPLDILIKTALTPHPKQAYGVTLPTSV 299  
 233 VLLAGHDNTLA-LGG-LDL-WOLP--QPDETPPGGALVPELWNR-----YVRVYQTM 280

Db 300 LFIAGHDTNLALGGALELNTLPGQDNTPPGGELVFERWRRLSDNSQWISLVFQTL 359  
Qy 281 DQLRNLEPLPRILP-----IPGC---GSEACSLSDPARLV---APAC 318  
Db 360 QQVRDRTPLSLNTPPGGVKLTLAGCERNAQGWCSLAGFTQIVNEARIPAC 410

## RESULT 10

ABR82310  
ID ABR82310 standard; protein; 412 AA.

XX ABR82310;  
XX 06-NOV-2003 (first entry)

XX Nov9X phytase polypeptide.

XX Nov9X; thermotolerant; phytase; phosphate; animal food; phosphorous;  
KW enzyme.

XX Synthetic.

XX Key Location/Qualifiers  
FH Misc-difference 1 /note= "encoded by GAG"

FT W02003057247-A1.

FT 17-JUL-2003.

PN 30-DEC-2002; 2002WO-US041785.

PD 28-DEC-2001; 2001US-0344523P.

XX (SYGN ) SYNGENTA PARTICIPATIONS AG.

XX (DIVE-) DIVERSA CORP.

XX Lanahan ML, Koepf E, Kretz K;

XX WPI; 2003-598328/56.

XX N-PSDB; ACF45789.

XX Preparing thermotolerant phytase useful for preparing animal feed or  
PT human food, by expressing in microbial host cell expression cassette  
PT comprising a promoter operably linked to a nucleic acid encoding the  
PT enzyme.

PS Claim 11; Fig 1B; 79pp; English.

XX The invention relates to preparing a thermotolerant phytase (I),  
CC involving expressing in a microbial host cell an expression cassette  
CC comprising a promoter operably linked to a nucleic acid molecule encoding  
CC (I). The thermotolerant phytase has a specific activity of greater than  
CC 400 U/mg, 600 U/mg or 800 U/mg at pH 4.5 and 37plusoc. The nucleic acid  
CC molecule encodes a fusion polypeptide comprising the thermotolerant  
CC phytase and a signal sequence which is operably linked to the  
CC thermotolerant phytase. The thermotolerant phytase is glycosylated and  
CC has a simulated gastric half-life of greater than 25 minutes at a pH  
CC greater than 2.0 and less than 4.0. (I) is useful for preparing animal  
CC feed, improving nutritive value of animal feed and human food, and for  
CC preparing human food. A heat-treated animal feed mixture comprising an  
CC inorganic phosphate at below 0.45% and (I) is useful for reducing feed  
CC conversion ratios or increasing weight gain of animals fed diets with  
CC inorganic phosphate at levels below 0.45%. It is also useful for  
CC decreasing phosphate levels in excreta from an animal. An animal feed  
CC composition comprising (I) is useful for decreasing the feed conversion  
CC ratio and increasing the weight gain of an animal. The composition is  
CC useful for minimizing inorganic dietary requirements of phosphorous in an  
CC animal, for enhancing utilization of phosphorous present in feed for an  
CC animal, and for increasing organic phosphate utilization from organic  
CC phosphorus sources in feed for an animal, where the feed increases the  
CC bioavailability of inorganic phosphorus in the feed to the animal. It is

CC also useful for decreasing phosphate levels in excreta from an animal. In  
CC all the above mentioned conditions, the feed is a poultry or swine feed.  
CC The thermotolerant phytase has a half-life of about 30 minutes in the  
CC digestive tract of the animal. The present sequence represents a  
CC thermotolerant Nov9X phytase

SQ Sequence 412 AA;

Query Match 37.4%; Score 616.5; DB 7; Length 412;

Best Local Similarity 45.5%; Pred. No. 4.2e-52;  
Matches 187; Conservative 41; Mismatches 70; Indels 113; Gaps 34;

Qy 19 AAAP-LKLEKWLTSRHGVRSPKA--LM--VSPD-WP-WPV---GLTPRGAALV-LLG 67  
Db 2 AQSEPELKLSESVIVSRHGVRAPTKATQLMQDTPDAMPVFKGLGELTRGGELIAYLG 61

Qy 68 Y-----RYFARGLLP--GCPAAG--TILADVDERTR-TGOAFAAGLAPGCAI-VH-----A 112  
Db 62 HYWRQLVADGLLPKCGCQSGQVAITADVDERTRKTGEAFAAGLAPDCAITVHTQADTS 121

Qy 113 GDDPIFHGLDTG-C-LDOA--DAILAAGGGGLTARHG-----LTAKVLNF--ASACL 159  
Db 122 SPDFLFNPLKTVGCOLDNANVTDAILERAGSGTADFTGHYQTAFRELERVLNFPQSNCL 181

Qy 160 -----ECARVVGRLGPLLASTISEIFLLEFLLEYAQGPMEVWGRI 196  
Db 182 KREKQDESCSLTQALPSELKVSADCVSLTGAVS--LASMLTEIFLLQQAQGMPEPGWRI 239

Qy 197 --SAEW---LLHNAQF-LANRTPIA--RATPI--LIVTALS-----PARV 232  
Db 240 TDSQWNTLSLHNAQFDLQRTPEVARSRATPLDLIKTALTTPHPKQKQAYGVTLTSTV 299

Qy 233 VLLAGHDTNLA-LGG-LDL-WQLP-QPETPPGGALVFELWNR-----YVRVMYQTM 280  
Db 300 LFIAGHDTNLALGGALELNTLPGQDNTPPGGELVFERWRRLSDNSQWISLVFQTL 359

Qy 281 DQLRNLEPLPRILP-----IPGC---GSEACSLSDPARLV---APAC 318  
Db 360 QQVRDRTPLSLNTPPGGVKLTLAGCERNAQGWCSLAGFTQIVNEARIPAC 410

## RESULT 11

ABP51937

ID ABP51937 standard; protein; 410 AA.

AC ABP51937;

DT 08-OCT-2002 (first entry)

DE Phytase protein sequence.

XX Phytase; enzyme; E.C.3.1.3.8; phytate; foodstuff; feed.

XX Unidentified.

XX WO200248332-A2.

XX 20-JUN-2002.

XX 12-DEC-2001; 2001WO-US048774.

XX 12-DEC-2000; 2000US-0255090P.

XX (DIVE-) DIVERSA CORP.

XX Short J, Mathur EJ, Richardson T, Robertson D, Barton N;

XX WPI; 2002-583504/62.

XX Novel recombinant phytase protein and polynucleotide for improving  
PT nutritional value of phytate-containing foodstuff, in animal feed and  
PT feed supplements and to degrade excess phytase from environment or  
PT sample.

```

XX PS Disclosure; Fig 9; 208pp; English.
XX CC
XX CC The present invention describes an isolated phytase protein (I). (I) can
XX CC be used for improving the nutritional value of a phytate-containing
XX CC foodstuff, by contacting the phytate-containing foodstuff with (I), where
XX CC the enzyme catalyses the liberation of inorganic phosphate from the
XX CC phytate-containing foodstuff, and so improving the nutritive value of the
XX CC contacted foodstuff. The liberation of inorganic phosphate occurs prior
XX CC to or after the ingestion of phytate-containing foodstuff by a recipient
XX CC organism. Nucleotide sequences (II) encoding (I) can be used for
XX CC producing an animal feed, by transforming a plant, plant portion or plant
XX CC cell with a nucleic acid expression vector, comprising (II), culturing
XX CC the plant, plant portion or plant cell under conditions in which the
XX CC phytase protein is expressed and converting the plant, plant portion or
XX CC plant cell into a composition suitable for animal feed. The animal is
XX CC preferably a monogastric animal or a ruminant. (I) and (II) are useful in
XX CC animal feed and feed supplements as well as in treatments to degrade or
XX CC remove excess phytate from the environment or a sample. (I) reduces
XX CC phytate levels in animal manure and so reduces the phosphate pollution of
XX CC the environment. The present sequence represents a phytase from the
XX CC present invention
XX CC
XX CC Sequence 410 AA;
XX CC
XX CC Query Match 37.2%; Score 613.5; DB 5; Length 410;
XX CC Best Local Similarity 45.5%; Pred. No. 8.3e-52;
XX CC Matches 186; Conservative 41; Mismatches 69; Indels 113; Gaps 34;
XX CC
XX CC 21 AEP-LKLEKVVILSRHGVRSPTKA--LM--VSPD-WP-WPV--GLTPRGAALV-LILGY- 68
XX CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX CC 2 SEPELKLESVVIVSRHGVRAPTKATQLMQDVTDPANWTPVKLGELTPRGGELIAYLGHY 61
XX CC
XX CC 69 ---RYFARGLLP--GCPAAG--TTLADVDERTR-TGQAFAGLAPGCAI-VH-----AGD 114
XX CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX CC 62 WRQRLVADGLLPKGCPCQSQGVAILIADVDERTRKTGEAFAGLAPDCAITVHTQADTSSP 121
XX CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX CC 115 DPFHGLDTG-C-LDQA---DAILAAGEGGTARHG-----LTLAKVLNF--ASACL-- 159
XX CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX CC 122 DPLFNPLKTVGCLQDNANVTDAILLERAGGSIAFTGHYQTAPELERVLNFPQSNCLKR 181
XX CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX CC 160 -----ECARVVGRGLPFIASLTSEIFLLEYAQGMPEVWGRI-- 196
XX CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX CC 182 EKODESCSLTQALPSELKVSADCVSLTGAVS--LASMLTEIFLLQQAGMPEPGWGRITD 239
XX CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX CC 197 SAEW-----LLHNAQF-LMNRTPYIA--RATPI--LIVTALS 234
XX CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX CC 240 SHQWNTLLSLHNAQFDLQRTPEVARSRATPLDLIKTALTPHPQKQAGVTLPTSVLF 299
XX CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX CC 235 LAGHDTNLA-LGG-LDL-WQLP-QDETTPPGGALVLFELNLR-----YVRVYQTMQ 282
XX CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX CC 300 IAGHDTNLANGGALELNWTLPGQPDNTPPGGELVFERMRRLSDNSQWIVQVSLVFTLQQ 359
XX CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX CC 283 LRNLEPLLPRILP-----IPGC---GSEAACSLSDFARLV----APAC 318
XX CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX CC 360 MRDKTPLUSLNTPPGEVKVTLTAGCERNAGMCSLAGFTQIVNEARIPAC 408
XX CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX CC
XX CC RESULT 12
XX CC AAB36257
XX CC ID AAB36257 standard; protein; 432 AA.
XX CC AC AAB36257;
XX CC
XX CC 12-SEP-2003 (revised)
XX CC DT 20-FEB-2001 (first entry)
XX CC
XX CC Lama2/APPA plasmid translated sequence.
XX CC
XX CC Transgenic animal; salivary protein; phytase; phosphorus; animal growth;
XX CC environmental pollution; pig.
XX CC
XX CC Mus musculus.

```

```

OS Escherichia coli.
XX Chimeric.
XX PN WO200064247-A1.
XX XX 02-NOV-2000.
XX PD
XX PF 20-APR-2000; 2000WO-CA000430.
XX PR 23-APR-1999; 99US-0130508P.
XX XX (UYGU-) UNIV GUELPH.
XX PA
XX PI Forsberg CW, Golovan S, Phillips JP;
XX XX WPI; 2000-687245/67.
XX DR N-PSDB; AAC68294.
XX DR
XX PT Transgenic non-human animal for gastrointestinal tract specific
XX PT expression of a protein, preferably phytase, comprises a nucleic acid
XX PT sequence including a heterologous transgene construct encoding the
XX PT protein.
XX PS Disclosure; Fig 5; 152pp; English.
XX CC
XX CC The present invention provides transgenic animals which produce desired
XX CC proteins, in this case pigs which expresses phytase in the salivary
XX CC gland. Low phytase production levels result in phytate in the diet being
XX CC excreted and causing phosphorus contamination in water, as well as
XX CC reducing the growth of animals. The invention provides a number of
XX CC transgenes containing the E. coli APPA phytase coding sequence. (Updated
XX CC on 12-SEP-2003 to standardise OS field)
XX CC
XX CC Sequence 432 AA;
XX CC
XX CC Query Match 37.2%; Score 613.5; DB 3; Length 432;
XX CC Best Local Similarity 45.3%; Pred. No. 8.9e-52;
XX CC Matches 194; Conservative 44; Mismatches 79; Indels 111; Gaps 36;
XX CC
XX CC 2 SILLAAALAAALAP--AAAAAEP-LKLEKVVILSRHGVRSPTKA--LM--VSPD-WP-W 52
XX CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX CC 3 ATLIPFLSLLIPTTQSAFAOSEPELKLESVVIVSRHGVRAPTKATQLMQDVTDPAMPTW 62
XX CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX CC 53 PV---GLTPRGAALV-LILGY-----RYFARGLL--PGCPAAG--TTLADVDERTR-TGQAF 99
XX CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX CC 63 PVKLGLWTPRGGELIAYLGHYQRLVADGLLAKGCPQSQGVAILIADVDERTRKTGEAF 122
XX CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX CC 100 AAGLAPGCAI-VH-----AGDDPIFHGLDTG-C-LDQA---DAILAAGEGGT--ARHGL 146
XX CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX CC 123 AAGLAPDCAITVHTQADTSSPDPLENPLKTVGCLQDNANVTDAILLRAGGSIAFTGHQ 182
XX CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX CC 147 T---LAKVLNF--ASACLE-----CARVVGRGLP--LASTLSEI 179
XX CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX CC 183 TAFRELERVLNFPQSNCLKREKQDESCSLTQALPSELKVSADNVSLTGAVSLASMLTEI 242
XX CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX CC 180 FLLLEYAQGMPEVWGRI--SAEW-----LLHNAQF-LMNRTPYIA--RATPI--LIVTALS 228
XX CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX CC 243 FLLQQAQGMPEPGWGRITDHSQWNTLLSLHNAQFLLQRTPEVARSRATPLDLIKTAL 302
XX CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX CC 229 -----PARVVLIGHDTNLA-LGG-LDL-WQLP-QDETTPPGGALVLFELNLR 271
XX CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX CC 303 PHPPKQAGVTLPTSVLFIAGHDTNLANGGALELNWTLPGQPDNTPPGGELVFERMR 362
XX CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX CC 272 -----YVRVYQTMQRLNLEPLLPRILP-----IPGC---GSEAACSLSDFARLV 314
XX CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX CC 363 LSDNSQWIVQVSLVFTLQQRDKTFLSLNTPGEVKLTLAGCERNAGMCSLAGFTQIV 422
XX CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX CC
XX CC 315 ----APAC 318
XX CC
XX CC 423 NEARIPAC 430
XX CC
XX CC RESULT 13

```



QY	229	-----PARVVLLAGHTNLA-IGG-LDL-WQLP-QDETTPGGALVFLWNR	271
Db	303	PHPPQKQAYGVTLSVLFIAGHDTNLANLGGALELNTLPGQPDNTPPGGELVFERWRR	362
QY	272	-----YVRVMYQTMQLENLSPLLPRILP-----IPGC---GSEAAACSLSDFARLV	314
Db	363	LSDNSQWITQVSLVFQTLQQRDKTFLSLNTPPGGVKUTLAGCBERNAGQWCSLAGFTQIV	422
QY	315	----APAC 318	
Db	423	NEARIPAC 430	
RESULT 14			
ID	AAB36262		
AD	AAB36262 standard; protein; 432 AA.		
XX	AAB36262;		
AC	AAB36262;		
DT	12-SEP-2003 (revised)		
DT	20-FEB-2001 (first entry)		
XX	SV40/APPA plasmid translated sequence.		
DE	Transgenic animal; salivary protein; phytase; phosphorus; animal growth; environmental pollution; pig.		
KW	Simian virus 40.		
OS	Escherichia coli.		
OS	Chimeric.		
XX	WO200064247-A1.		
PN	02-NOV-2000.		
PD	20-APR-2000; 2000WO-CA000430.		
XX	23-APR-1999; 99US-0130508P.		
PR	(UYGU-) UNIV GUELPH.		
PA	Foreberg CW, Golovan S, Phillips JP;		
XX	WPI; 2000-687245/67.		
XX	N-PSDB; AAC68299.		
PT	Transgenic non-human animal for gastrointestinal tract specific expression of a protein, preferably phytase, comprises a nucleic acid sequence including a heterologous transgene construct encoding the protein.		
PT	Disclosure; Fig 22; 152pp; English.		
PS	The present invention provides transgenic animals which produce desired proteins, in this case pigs which expresses phytase in the salivary gland. Low phytase production levels result in phytate in the diet being excreted and causing phosphorus contamination in water, as well as reducing the growth of animals. The invention provides a number of transgenes containing the E. coli APPA phytase coding sequence. (Updated on 12-SEP-2003 to standardise OS field)		
XX	Sequence 432 AA;		
CC	Query Match 37.2%; Score 613.5; DB 3; Length 432;		
CC	Best Local Similarity 45.3%; Pred. No. 8.9e-52;		
CC	Matches 194; Conservative 44; Mismatches 79; Indels 111; Gaps 3		
QY	2	SLTLAALLAALAP--AAAAEP-LKLEKVVILSRHGVRSPTKA--LM--VSPD-WP-W 52	
Db	3	ALILFPLSLILPLTQSPAPQSEFELKLESVVVSRHGVRAPTATQOLMQDVTEDANPTW 62	
QY	53	PV---GLTPRGAALV-LIGY-----RYFARGLL--PGCPAAG--TILADVDERTR-TQAAF 99	

Db 63 PVKLGWLTFRGGELIAYGHYQORLVADGLLAKKGCPSQGVVAIIADVDERTKGTGEAF 122  
 QY 100 AAGLAPGCAI-VH-----AGDDPIFHGLDTG-C-LDOA-----DAILAAGEGLT--ARHGL 146  
 Db 123 AAGLAPGCAI-VH-----AGDDPIFHGLDTG-C-LDOA-----DAILAAGEGLT--ARHGL 146  
 QY 147 T-----LAKVNF--ASACLE-----CARVVGRLGGL--LASTLSEI 179  
 Db 183 TAFRELERVNFPPQSNLCKREKQDESCSLTQALPSELKVSADNVSLTGAVSLASMLTEI 242  
 QY 180 FLLEYAOGPMVEVWGRI--SAEW-----LHNAQF-LMNRTPYA--RATPI--LIVTALS 228  
 Db 243 FLLOQAQMPGPGWGRITDHSQWNTLLSLHNAQFYLLQRTPEVARSRAATPLDLIKTALT 302  
 QY 229 -----PARVVLLAGHDTNLA-LGG-LDL-WOLP-OPDETTPPGGALVFELWNR 271  
 Db 303 PHPPQKQAYGVTLFTSVLFIAGHDTNLANLGGALELNTLPGQPDNTPPGGELVFERWR 362  
 QY 272 -----YVRVMYQTMQDLNLEPLPRILP-----IPGC-----GSEAAACSLSDFA 314  
 Db 363 LSDNSQWISLVFQTLQOMRDKTFLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIV 422  
 QY 315 ----APAC 318  
 Db 423 NEARIPAC 430

RESULT 15

AAB36259  
 ID AAB36259 standard; protein; 432 AA.

AC AAB36259;

DT 12-SEP-2003 (revised)

DT 20-FEB-2001 (first entry)

XX R15/APPA plasmid translated sequence.

XX Transgenic animal; salivary protein; phytase; phosphorus; animal growth;  
 KW environmental pollution; pig.

XX Rattus sp.

OS Escherichia coli.

OS Chimeric.

PN WO200064247-A1.

XX 02-NOV-2000.

XX 20-APR-2000; 2000WO-CA000430.

XX 23-APR-1999; 99US-0130508P.

XX (UYGU-) UNIV GUELPH.

XX Forsberg CW, Golovan S, Phillips JP;

XX WPI; 2000-687245/67.

XX N-PSDB; AAC68296.

XX Transgenic non-human animal for gastrointestinal tract specific  
 PT expression of a protein, preferably phytase, comprises a nucleic acid  
 PT sequence including a heterologous transgene construct encoding the  
 PT protein.  
 XX Disclosure; Fig 19; 152pp; English.

XX The present invention provides transgenic animals which produce desired  
 CC proteins, in this case pigs which expresses phytase in the salivary  
 CC gland. Low phytase production levels result in phytate in the diet being  
 CC excreted and causing phosphorus contamination in water, as well as  
 CC reducing the growth of animals. The invention provides a number of  
 CC transgenes containing the E. coli APPA phytase coding sequence. (Updated

CC on 12-SEP-2003 to standardise OS field)

XX Sequence 432 AA;

Query Match 37.2%; Score 613.5; DB 3; Length 432;

Best Local Similarity 45.3%; Pred. No. 8.9e-52;

Matches 194; Conservative 44; Mismatches 79; Indels 111; Gaps 36;

QY 2 SLLLAALLAALAP--AAAAAEP-LKLEKVVILSRHGVRSPTKA--LM--VSPD-WP-W 52

Db 3 AILIFPLSLLPLTPQSAFAQSEPELKLESVVIVSRHGVRAPTKATQLMQDVTDPDAPTW 62

QY 53 PV---GLTPRGAAIV-LLGY-----RYFARGLL--PGCPAAG--TILADVDERTR-TGQAF 99

Db 63 PVKLGWLTFRGGELIAYGHYQORLVADGLLAKKGCPSQGVVAIIADVDERTKGTGEAF 122

QY 100 AAGLAPGCAI-VH-----AGDDPIFHGLDTG-C-LDOA-----DAILAAGEGLT--ARHGL 146

Db 123 AAGLAPGCAI-VH-----AGDDPIFHGLDTG-C-LDOA-----DAILAAGEGLT--ARHGL 146

QY 147 T-----LAKVNF--ASACLE-----CARVVGRLGGL--LASTLSEI 179

Db 183 TAFRELERVNFPPQSNLCKREKQDESCSLTQALPSELKVSADNVSLTGAVSLASMLTEI 242

QY 180 FLLEYAOGPMVEVWGRI--SAEW-----LHNAQF-LMNRTPYA--RATPI--LIVTALS 228

Db 243 FLLOQAQMPGPGWGRITDHSQWNTLLSLHNAQFYLLQRTPEVARSRAATPLDLIKTALT 302

QY 229 -----PARVVLLAGHDTNLA-LGG-LDL-WOLP-OPDETTPPGGALVFELWNR 271

Db 303 PHPPQKQAYGVTLFTSVLFIAGHDTNLANLGGALELNTLPGQPDNTPPGGELVFERWR 362

QY 272 -----YVRVMYQTMQDLNLEPLPRILP-----IPGC-----GSEAAACSLSDFA 314

Db 363 LSDNSQWISLVFQTLQOMRDKTFLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIV 422

QY 315 ----APAC 318

Db 423 NEARIPAC 430

Search completed: April 30, 2004, 12:38:48  
 Job time : 42.7195 secs

GenCore version 5.1.6  
 Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 12:36:16 ; Search time 12.7422 Seconds  
 (without alignments)  
 1288.405 Million cell updates/sec

Title: US-10-021-723B-14  
 Perfect score: 1647  
 Sequence: 1 VSLLAALLAALAPAAAA.....CGSEACSLSDFAFLVAPAC 318

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: Issued Patents AA:\*
- 2: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	611.5	37.1	440	3	US-09-259-214-2
2	611.5	37.1	440	3	US-09-318-528-2
3	611.5	37.1	440	3	US-09-291-931-2
4	608.5	36.9	433	4	US-09-540-149A-1
5	556	33.8	423	2	US-08-910-798-2
6	272.5	16.5	421	4	US-09-489-039A-7512
7	265.5	16.1	522	4	US-09-489-039A-13501
8	116	7.0	798	4	US-09-252-991A-25309
9	103.5	6.3	386	1	US-08-758-213-1
10	103.5	6.3	386	2	US-08-692-787-48
11	103.5	6.3	386	3	US-09-097-199-48
12	103.5	6.3	514	4	US-09-252-991A-25845
13	103.5	6.3	515	2	US-09-146-283-2
14	103.5	6.3	515	3	US-08-579-823A-2
15	103.5	6.3	515	3	US-09-344-185-2
16	102.5	6.2	4551	3	US-09-320-878-1
17	102.5	6.2	4551	4	US-09-141-908-2
18	102.5	6.2	4551	4	US-09-657-440-1
19	102.5	6.2	4613	3	US-09-105-537-31
20	102.5	6.2	11877	3	US-09-105-537-6
21	99	6.0	638	4	US-09-252-991A-18036
22	96.5	5.9	465	4	US-09-044-718-79
23	93.5	5.7	318	3	US-08-680-506-3
24	93.5	5.7	449	3	US-08-680-506-7
25	93.5	5.7	831	4	US-09-252-991A-30097
26	92.5	5.6	281	4	US-09-252-991A-19858
27	92.5	5.6	467	4	US-09-684-855-165

28 91.5 5.6 439 4 US-09-044-718-3

29 91.5 5.6 440 4 US-09-684-855-106

30 91.5 5.6 440 4 US-09-684-855-128

31 91.5 5.6 440 4 US-09-684-855-151

32 91.5 5.6 449 4 US-09-044-718-12

33 91.5 5.6 465 3 US-08-868-435-33

34 91.5 5.6 465 4 US-08-744-231-33

35 91.5 5.6 465 4 US-09-044-718-78

36 91.5 5.6 465 4 US-09-636-499-6

37 91.5 5.6 465 4 US-09-273-871A-8

38 91.5 5.6 499 4 US-09-489-039A-9208

39 91.5 5.6 560 4 US-09-252-991A-28159

40 91.5 5.6 579 4 US-09-252-991A-18063

41 91 5.5 580 4 US-09-252-991A-27245

42 91 5.5 841 4 US-09-252-991A-28589

43 91 5.5 1002 4 US-09-252-991A-19662

44 90.5 5.5 461 4 US-09-252-991A-19438

45 90 5.5 581 4 US-09-252-991A-17734

#### ALIGNMENTS

##### RESULT 1

US-09-259-214-2

; Sequence 2, Application US/09259214A

; Patent No. 6110719

; GENERAL INFORMATION:

; APPLICANT: Kretz, Keith

; TITLE OF INVENTION: NOVEL PHYTASE

; FILE REFERENCE: DIVER1370-1

; CURRENT APPLICATION NUMBER: US/09/259,214A

; CURRENT FILING DATE: 1999-03-01

; EARLIER APPLICATION NUMBER: 08/910,798

; EARLIER FILING DATE: 1997-08-13

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: FASTSEQ for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 440

; TYPE: PRT

; ORGANISM: Escherichia coli

; US-09-259-214-2

Query Match 37.1%; Score 611.5; DB 3; Length 440;  
 Best Local Similarity 45.1%; Pred. No. 1.5e-59;  
 Matches 193; Conservative 45; Mismatches 79; Indels 111; Gaps 36;

QY	2	SLLAALLAALAP--AAAAEP-LKLEKVVLSRHGVRSPKA--LM--YSPD-WP-W	52
DB	3	ALIPPLSLILPLTPQSAFAQSEPELKLESVVIVSRHGVRATKATQTLQMDVTPDAWPTW	62
QY	53	PV---GLTPRGARLV-LLGY-----RYFARGLL--PCPPAAG--TILADVDETR-TGOAF	99
DB	63	PVKLGWLTFRGGELIAYLGHYQRLVADGLLAKKGCPOSGQVAILADVDETRKTGEAF	122
QY	100	AAGLAPGCAI-VH-----AGDDPIPHGLDTG-C-LDOA---DAILAAGSGGLT--ARHGL	146
DB	123	AAGLAPDCAITVHTQADTSSPDPLFNPLKTVGVCQLDNANVTAILSRAGGSIAFTGCHQ	182
QY	147	T-----LAKVINF--ASACLE-----CARVVGRGLGL-LASTLSI	179
DB	183	TAFRELERVLNFPQSNLCRKQDSSCSLTQALPSELKVSADNVSLTGAVSLASMLTEI	242
QY	180	FLLEYAQGPMEVWGRI--SAEW-----LIHNQF-LMNETPYIA--RATPI--LIVTALS	228
DB	243	FLIQAQGPMEPGWGRTDTSQWNTLSLHNAQVILLQRTPEVARSRATPLDLIIVALT	302
QY	229	-----PARVLLAGHTNLA-IGGG-LDL-WOLP-OPDETTPFGALVFLWNR	271
DB	303	PHFPQKAYGVLTPTSVLFIAGHTNLANLGGALELNWTLFGQPDNTPFGGLVFERWR	362
QY	272	-----YRVVYQNDQLRNLEPLRLPLP-----IPGC---GSEACSLSDFAFLV	314

Db 363 LSDNSQWIOVSLVFQTLQMRDKTPLSLNTPPGEVKLTLAGCBNAQMGCSLAGFTQIV 422  
 Qy 315 ----APAC 318  
 Db 423 NEARIPAC 430

RESULT 2

US-09-318-528-2  
 ; Sequence 2, Application US/09318528  
 ; Patent No. 6183740  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kretz, Keith  
 ; TITLE OF INVENTION: NOVEL PHYTASE  
 ; FILE REFERENCE: 09010/029003  
 ; CURRENT APPLICATION NUMBER: US/09/318,528  
 ; CURRENT FILING DATE: 1999-05-25  
 ; EARLIER APPLICATION NUMBER: 09/291,931  
 ; EARLIER FILING DATE: 1999-04-13  
 ; EARLIER APPLICATION NUMBER: 08/910,798  
 ; EARLIER FILING DATE: 1997-08-13  
 ; EARLIER APPLICATION NUMBER: 09/259,214  
 ; EARLIER FILING DATE: 1999-03-01  
 ; NUMBER OF SEQ ID NOS: 4  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 2  
 ; LENGTH: 440  
 ; TYPE: PRT  
 ; ORGANISM: Escherichia coli  
 ; US-09-318-528-2

Query Match 37.1%; Score 611.5; DB 3; Length 440;  
 Best Local Similarity 45.1%; Pred. No. 1.5e-59;  
 Matches 193; Conservative 45; Mismatches 79; Indels 111; Gaps 36;

Qy 2 SLLAALLAALAP--AAAAEP--LKLEKVVILSRHGVSPKTA--LM--VSPD-WP-W 52  
 Db 3 AILIPFLSLIPLTPQSAFAQSBPELKESVIVSRHGVRAPTKATQMLQMDVTPDAWPTW 62  
 Qy 53 PV---GLTPRGAALV--LLGY---RYFARGLL--PGCPAAG--TILADVDERT--TGOAF 99  
 Db 63 PVKLGWLTTPRGGLIAYLGHVQRLVADGLLAKKGCPOSGQVAILADVDERTKTGEAF 122  
 Qy 100 AAGLAPGCAI-VH-----AGDDPIFHGLDTG-C-LDOA---DAILAAGEGLT--ARHGL 146  
 Db 123 AAGLAPDCAITVHTQADTSSPDLFNLKTVGVQOLDNANVTDAILSRAGGSIAFTGHRQ 182  
 Qy 147 T---LAKVLFN--ASACLE-----CARVVGRLGFL--LASTLSEI 179  
 Db 183 TAFRELERVLNFPQSNCLKKEKQDESCSLTQALPSELKVSADNVSLTGAVSLASMLTEI 242  
 Qy 180 FLLEYAQGPMEVGVGRI--SAEW-----LLHNAOF--LMNTPPYIA--PATPI--LIVTALS 228  
 Db 243 FLQQAQGMPEPGWGRITDHSQWNTLLSLHNAQFYLLQRTPEVARSEATPLDLIMAALT 302  
 Qy 229 -----PARVLLAGHDNTLA-LGG-LDL-WQLP-OPDETTPGGALVPELWNR 271  
 Db 303 PHPPQKAYGVTLPTSVLFAGHDTNLNAGGLELNLWTLPGQPDNTPPGGELVFERWR 362  
 Qy 272 -----YRVVMYQTMQDLRNLEPLLRILP-----IPGC---GSEAACSLSDPARLV 314  
 Db 363 LSDNSQWIOVSLVFQTLQMRDKTPLSLNTPPGEVKLTLAGCBNAQMGCSLAGFTQIV 422  
 Qy 315 ----APAC 318  
 Db 423 NEARIPAC 430

RESULT 3

US-09-291-931-2  
 ; Sequence 2, Application US/09291931A  
 ; Patent No. 6190897  
 ; GENERAL INFORMATION:

; APPLICANT: Kretz, Keith  
 ; TITLE OF INVENTION: NOVEL PHYTASE  
 ; FILE REFERENCE: 09010/029003  
 ; CURRENT APPLICATION NUMBER: US/09/291,931A  
 ; CURRENT FILING DATE: 1999-04-13  
 ; EARLIER APPLICATION NUMBER: 08/910,798  
 ; EARLIER FILING DATE: 1997-08-13  
 ; EARLIER APPLICATION NUMBER: 09/259,214  
 ; EARLIER FILING DATE: 1999-03-01  
 ; NUMBER OF SEQ ID NOS: 4  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 2  
 ; LENGTH: 440  
 ; TYPE: PRT  
 ; ORGANISM: Escherichia coli  
 ; US-09-291-931-2

Query Match 37.1%; Score 611.5; DB 3; Length 440;  
 Best Local Similarity 45.1%; Pred. No. 1.5e-59;  
 Matches 193; Conservative 45; Mismatches 79; Indels 111; Gaps 36;

Qy 2 SLLAALLAALAP--AAAAEP--LKLEKVVILSRHGVSPKTA--LM--VSPD-WP-W 52  
 Db 3 AILIPFLSLIPLTPQSAFAQSBPELKESVIVSRHGVRAPTKATQMLQMDVTPDAWPTW 62  
 Qy 53 PV---GLTPRGAALV--LLGY---RYFARGLL--PGCPAAG--TILADVDERT--TGOAF 99  
 Db 63 PVKLGWLTTPRGGLIAYLGHVQRLVADGLLAKKGCPOSGQVAILADVDERTKTGEAF 122  
 Qy 100 AAGLAPGCAI-VH-----AGDDPIFHGLDTG-C-LDOA---DAILAAGEGLT--ARHGL 146  
 Db 123 AAGLAPDCAITVHTQADTSSPDLFNLKTVGVQOLDNANVTDAILSRAGGSIAFTGHRQ 182  
 Qy 147 T---LAKVLFN--ASACLE-----CARVVGRLGFL--LASTLSEI 179  
 Db 183 TAFRELERVLNFPQSNCLKKEKQDESCSLTQALPSELKVSADNVSLTGAVSLASMLTEI 242  
 Qy 180 FLLEYAQGPMEVGVGRI--SAEW-----LLHNAOF--LMNTPPYIA--PATPI--LIVTALS 228  
 Db 243 FLQQAQGMPEPGWGRITDHSQWNTLLSLHNAQFYLLQRTPEVARSEATPLDLIMAALT 302  
 Qy 229 -----PARVLLAGHDNTLA-LGG-LDL-WQLP-OPDETTPGGALVPELWNR 271  
 Db 303 PHPPQKAYGVTLPTSVLFAGHDTNLNAGGLELNLWTLPGQPDNTPPGGELVFERWR 362  
 Qy 272 -----YRVVMYQTMQDLRNLEPLLRILP-----IPGC---GSEAACSLSDPARLV 314  
 Db 363 LSDNSQWIOVSLVFQTLQMRDKTPLSLNTPPGEVKLTLAGCBNAQMGCSLAGFTQIV 422  
 Qy 315 ----APAC 318  
 Db 423 NEARIPAC 430

RESULT 4

US-09-540-149A-1  
 ; Sequence 1, Application US/09540149A  
 ; Patent No. 6511699  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lei, Xingen  
 ; TITLE OF INVENTION: ENZYMES WITH IMPROVED PHYTASE ACTIVITY  
 ; FILE REFERENCE: 19603/2791  
 ; CURRENT APPLICATION NUMBER: US/09/540,149A  
 ; CURRENT FILING DATE: 2000-03-31  
 ; PRIOR APPLICATION NUMBER: 60/127,032  
 ; PRIOR FILING DATE: 1999-03-31  
 ; NUMBER OF SEQ ID NOS: 9  
 ; SOFTWARE: Patent in Ver. 2.1  
 ; SEQ ID NO 1  
 ; LENGTH: 433  
 ; TYPE: PRT  
 ; ORGANISM: Escherichia coli  
 ; FEATURE:

NAME/KEY: UNSURE  
LOCATION: (433)  
OTHER INFORMATION: Xaa at position 433 in this sequence is unknown  
US-09-540-149A-1

Query Match 36.9%; Score 608.5; DB 4; Length 433;  
Best Local Similarity 45.1%; Pred. No. 3.1e-59;  
Matches 193; Conservative 43; Mismatches 81; Indels 111; Gaps 36;

QY 2 SLLAALLAALAP--AAAAEP-LKLEKVVILSRHGVRSPTKA--LM--VSPD-WP-W 52  
DB 3 AILIPFLSLIPLTPQSAFAQSEPELKESVIVSRHGVRAPTKATQLMQDVTDPANPTW 62  
QY 53 PV---GLTPGGAALV-LLGY---RYFARGLL--PGCPAAG--TILADVDETR-TGOAF 99  
DB 63 PVKLGWLTFRGGELIAYLGHYQORLVADGLLAKKGCPOGQVAILADVDETRTKTGEAF 122  
QY 100 AAGLAPGCAI-VH-----AGDDPIFRGLDTG-C-LDOA---DAILAAGEGGLT--ARHGL 146  
DB 123 AAGLAPDCAITVHTQADTSPPDLFNLKTVGVCOLDNANVTDAILSRAGGSIAADFTGHRQ 182  
QY 147 T-----LAKVLNEA--SACLE-----CARVGRGLGFL--LASTLSEI 179  
DB 183 TAFRELVLNFSQNLNKLREKQDESCSLTQALPSELKVSADNVSILTGAVALSLMTET 242  
QY 180 FLLEYAQGPMEVGVGRI--SAEW-----LLHNAOF-LMRTPTPIA--RATPI--LIVTALS 228  
DB 243 FLLOQAQGMPEPCWGRITDQNTLSLHNAQFYLLQRTPEVARSRATPLDLMAALT 302  
QY 229 -----PARVLLAGHDNTLA-LGG-LDL-WQLP-OPDETTPGGALVVELWNR 271  
DB 303 PHPPQKQAVGVTUFTSVLFTAGHDNTLANLGGALELNTLPGQPDNTTPGGELVFERWR 362  
QY 272 -----YRVVYQTMQDLNLEPLLRILP-----IPGC---GSEAACSLSDFARLV 314  
DB 363 LSDNSQWIVSLVFTLQQRDKTSLNTPPGCVKLTLAGCERNAGQWCSLAGTQIV 422  
QY 315 -----APAC 318  
DB 423 NEARIPAC 430

## RESULT 5

US-08-910-798-2  
Sequence 2, Application US/08910798  
Patent No. 5875997  
GENERAL INFORMATION:  
APPLICANT: KRETTZ  
TITLE OF INVENTION: NOVEL PHYTASE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: California  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/910,798  
FILING DATE: August 13, 1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: HAILE, PH.D., Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 09010/029001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/678-5070  
TELEFAX: 619/678-5099

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 423 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-910-798-2

Query Match 33.8%; Score 556; DB 2; Length 423;

Best Local Similarity 43.1%; Pred. No. 2.2e-53;  
Matches 182; Conservative 44; Mismatches 80; Indels 116; Gaps 35;

QY 2 SLLAALLAALAP--AAAAEP-LKLEKVVILSRHGVRSPTKA--LM--VSPD-WP-W 52  
DB 3 AILIPFLSLIPLTPQSAFAQSEPELKESVIVSRHGVRAPTKATQLMQDVTDPANPTW 62  
QY 53 PV---GLTPGGAALV-LLGY---RYFARGLL--PGCPAAG--TILADVDETR-TGOAF 99  
DB 63 PVKLGWLTFRGGELIAYLGHYQORLVADGLLAKKGCPOGQVAILADVDETRTKTGEAF 122  
QY 100 AAGLAPGCAI-VHAGDDPIFRGLDTG-C-LDOA---DAILAAGEGGLT--ARHGLT---L 148  
DB 123 T-----SSPDLFNLKTVGVCOLDNANVTDAILSRAGGSIAADFTGHRQTAFREL 171  
QY 149 AKVLNF--SACLE-----CARVGRGLGFL--LASTLSEIFLLEYA 185  
DB 172 ERLNFPQSNLCKREKQDESCSLTQALPSELKVSADNVSILTGAVALSLMTETFLLOQA 231  
QY 186 QGPMEVGVGRI--SAEW-----LLHNAOF-LMRTPTPIA--RATPI--LIVTALS----- 228  
DB 232 QGMEPGMGRTDQNTLSLHNAQFYLLQRTPEVARSRATPLDLMAALTTPHPPOK 291  
QY 229 -----PARVLLAGHDNTLA-LGG-LDL-WQLP-OPDETTPGGALVVELWNR----- 271  
DB 292 QAYGVTUFTSVLFTAGHDNTLANLGGALELNTLPGQPDNTTPGGELVFERWRRLSDNSQ 351  
QY 272 --YRVVYQTMQDLNLEPLLRILP-----IPGC---GSEAACSLSDFARLV-----AP 316  
DB 352 WIQVSLVFTLQQRDKTSLNTPPGCVKLTLAGCERNAGQWCSLAGTQIVNEARIP 411  
QY 317 AC 318  
DB 412 AC 413

## RESULT 6

US-09-489-039A-7512  
Sequence 7512, Application US/09489039A  
Patent No. 6610836  
GENERAL INFORMATION:  
APPLICANT: Gary Breton et. al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
FILE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 2709.2004001  
CURRENT APPLICATION NUMBER: US/09/489,039A  
CURRENT FILING DATE: 2000-01-27  
PRIOR APPLICATION NUMBER: US 60/117,747  
PRIOR FILING DATE: 1999-01-29  
NUMBER OF SEQ ID NOS: 14342  
SEQ ID NO 7512  
LENGTH: 421  
TYPE: PRT  
ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-7512

Query Match 16.5%; Score 272.5; DB 4; Length 421;  
Best Local Similarity 28.0%; Pred. No. 9.6e-22;  
Matches 114; Conservative 62; Mismatches 130; Indels 101; Gaps 26;

QY 6 AAILAALAPAAAAEP-LKLEKVVILSRHGVRSPTKA--LM--VSPD-WP-W 55  
DB 19 AVMLSAGAAQDKAAPGEGYQLQVILMSRLNRLAPLANSVLEQSTAKAWPQWDVFGGQ 78

[illegible]

```

RESULT 7
US-09-489-039A-13501
; Sequence 13501 Application US/09489039A
; Patent No. 8610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 13501
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13501

```

```

Query Match      16.1%; Score 265.5; DB 4; Length 522;
Best Local Similarity 30.5%; Pred. No. 8e-21;
Matches 119; Conservative 44; Mismatches 142; Indels 85; Gaps 24;

Qy      3 LLLAALLAALAPAAAAEPLKLEKVVLSRHGVRSP-----KALMVSPDWPVPV--- 54
Db      112 LFIAALPPLLALQAAAA--DWQLEKVVLSRHGIRPPTAGNRBAIEAATGCRPTWETH 169
Qy      55 --GLTPRGAALVL-----LGRYFARGILL-PCGFAAGTIL--ADVDERTR-TGQAFAGL 103
Db      170 DGELTGHGYAAVNVKGRAGEGHYRQLGLLQAGCTABSIYVRASPLQTRATAQALVDGA 229
Qy      104 APGCAL-VH--AGD-DPIFHGLDTCGLDQAD-----AIIAGEBGLTARHGLTAKVLNF 154
Db      230 PFGCGVAIHYVSGDADPLFQ-TDKFAATQTFAPQLAAVKEKAGDLAQRQALAPTQLL 288
Qy      155 ASACLECARVV-----GRGLPLLASTLS-----BIFILEYAQG-PM-EVGNGR 195
Db      289 KOAVCOADKPCPIFTPTMQVQSGSKGKTTIGLSVMANNVETLRIGWSENIPLSQLAWGK 348
Qy      196 ISAEW-----LLHNAQFLMNRPTIYATAPLITVTAI-----SP-ARVILLAGHD 239
Db      349 LTAQEQITALLPLLTENYDLTNDVLYTAQKGSVLLNAMLGVKPEANPNVRWLLVAHD 408
Qy      240 TNIALGGJDL---WOLF--QPDETPPGGALVELM-----NRYVRVMYQT--MDQLRVL 286
Db      409 TNIAVRITMNPFWOLPGVSRGNIIPGSSVLIERWENAKSGERYLVRVFOAGLDDLRLR 468

```

QY      287 E-----PLLPRLPIPGCGSEACSLSDP 310  
         :|: |  ||       |:|  
Db      469 QTPDAQHFMLEOENHQPGRQTVDVGTLCFP 498

RESULT 8  
US-09-252-991A-25309  
; Sequence 25309, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: 1999-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 25309  
; LENGTH: 798  
; TYPE: PRF  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-25309

Query Match	7.0%; Score 116; DB 4; Length 798;
Best local Similarity	27.0%; Pred. No. 0.00074;
Matches	77; Conservative 118; Indels 66; Gaps 12;
QY	67 GY-----RYFARGLLPGCPAAGTILADVDSRTTG-----QAFPAAGLAPCAIVHA 112
Db	486 GYDVRERTWASDVLELVEFGRLAALPELISPGACIGNLRPEAAALGLAPHTRVACG 545
QY	113 GDDPIFHGLDTG-----CLDQADAILAAGEGGLTABHG-----LTIAKV 151
Db	546 GGDNMLAAIGTGNIRPGLLTASGTSTLSAYAPERLVSPHGEIATTCASGGWLPLACT 605
QY	152 INFASACLECARVVGRIGPLIASTLSIFILEYAQGPMEVGMWGSIASFWLLHNAQFLMNR 211
Db	606 MNLTGAC-----GLVQDILLHLDLDFSNLA-AQAP-----VGAEGILMLPFFDGER 650
QY	212 TPYTARATPLI-VTALS PARVVLLAGHDINLALG-----GLDLWQ-----LPQDETPPGCAL 264
Db	651 VPALPHASASLHGHTAANLGRANICRAVLGETATGLRYGLDLLRASGLPGEIRLVGAA 710
QY	265 VFELWNRVVRVWYQTMQRLNLEPLLPRILPIPGCGS--EAACSL 307
Db	711 KNDLWRR-----TLADLLGLPLVCPROTEAAALGAALQAWSL 748

RESULT 9  
US-08-758-213-1  
; Sequence 1, Application US/08758213  
; Patent No. 5763490  
; GENERAL INFORMATION:  
; APPLICANT: Lebioda, L.  
; APPLICANT: Jakob, C.  
; TITLE OF INVENTION: TREATING PROSTATE CANCER BY INHIBITING PROSTATIC ACID PHOSP  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Michael A. Mann, P.A.  
; STREET: Post Office Box 7908  
; CITY: Columbia  
; STATE: South Carolina  
; COUNTRY: United States  
; ZIP: 29202  
; COMPUTER READABLE FORM: On attached diskette  
; MEDIUM TYPE: Diskette 3.5 inch, 1.4 MB storage  
; COMPUTER: Apple Macintosh TM  
; OPERATING SYSTEM: Macintosh TM  
; SOFTWARE: Microsoft Word 5.0 TM

CITY: Houston  
STATE: Texas  
COUNTRY: United States of America  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
FILING DATE: 20-MAY-1996  
APPLICATION NUMBER: 08/309,091  
FILING DATE: 09/20/94  
ATTORNEY/AGENT INFORMATION:  
NAME: Mann, Michael A.  
REGISTRATION NUMBER: 32,825  
REFERENCE/DOCKET NUMBER: 96-2578  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (803) 254-8472  
TELEFAX: (803) 254-3760  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 386 amino acids  
TYPE: Amino acid  
STRANDEDNESS: No. 5763490 Applicable  
TOPOLOGY: Linear  
US-08-758-213-1

Query Match 6.3%; Score 103.5; DB 1; Length 386;  
Best Local Similarity 23.8%; Pred. No. 0.0062;  
Matches 92; Conservative 38; Mismatches 148; Indels 109; Gaps 20;

QY 2 SLIIAALLAALAPAAAAEPLKLVILSRHGVSPPTKALMVSP--DWPWPVG---L 56  
DB 13 SLISGLFLFLFFWLDORSVLAKELF--VTLVFRHGDRSPIDTPTDPIKSSWPQGFQGL 70  
QY 57 TPRG-----AALVILGYRYF-----ARGLLPGCPAAGTILADVDRTRTGOA 98  
DB 71 TQLGMEQHYELGEYIRKRYKFLNESYKHEQVYIRS-----TDVDRTLMSAMT 118  
QY 99 FAAGLAP--GCAI-----VHA-----GDD-----PIFHGLDTGCLDQ--- 128  
DB 119 NLALFPPEGVSINWPILLMQPIVHTVPLSEDLQLLYPFRNCPRFOLESETLKSEEFQ 178  
QY 129 -----ADAILAAGEGGLTARHGLTLAKVLNFAFASACLECARVVGRLGPIILAS----- 174  
DB 179 KRLHPYKDFIATLKG--LSGLHGQDLFGIWSKVYDPLCYESVHNFTLPSWATEDWTMKLR 236  
QY 175 TLSIFILLEYAOGPMEVWGGRISAEWLLHNAOFLMNRTPYIARATPILIVTALSAPARVL 234  
DB 237 ELSLSLLSLYGIHKQEKSRLOGGVLVNE---ILN---HMKRATQI-----PSYKKLIM 285  
QY 235 LAGHDTNLALGGLDLWQLPQDPETPPGGA-----LVFELMNRVYRVVYQTMQDLRNLEPL 289  
DB 286 YSAHDTTVS--GLQWALDVYNGLLPPYASCHLTLYFEKGEYFVEMY-----RNETQH 337  
QY 290 LPRILPIPGCGSEACSLSDFARLVAP 316  
DB 338 EYPVLMPLGC--SPSCPLERFAELVGP 362

RESULT 10  
US-08-692-787-48

; Sequence 48, Application US/08692787  
; Patent No. 582864  
; GENERAL INFORMATION:  
; APPLICANT: Ar, Gang  
; APPLICANT: O'Hara, S. Mark  
; APPLICANT: Veltri, Robert  
; APPLICANT: Veltri, Robert  
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,  
; TITLE OF INVENTION: PROGNOSIS AND MANAGEMENT OF PROSTATE  
; TITLE OF INVENTION: DISEASE  
; NUMBER OF SEQUENCES: 82  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P.O. Box 4433

CITY: Houston  
STATE: Texas  
COUNTRY: United States of America  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
FILING DATE: 09/20/94  
APPLICATION NUMBER: 08/309,091  
FILING DATE: 20-MAY-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Mann, Michael A.  
REGISTRATION NUMBER: 32,825  
REFERENCE/DOCKET NUMBER: 96-2578  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (803) 254-8472  
TELEFAX: (803) 254-3760  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 386 amino acids  
TYPE: Amino acid  
TOPOLOGY: Linear  
MOLSCULE TYPE: Protein  
US-08-692-787-48

Query Match 6.3%; Score 103.5; DB 2; Length 386;  
Best Local Similarity 23.8%; Pred. No. 0.0062;  
Matches 92; Conservative 38; Mismatches 148; Indels 109; Gaps 20;

QY 2 SLIIAALLAALAPAAAAEPLKLVILSRHGVSPPTKALMVSP--DWPWPVG---L 56  
DB 13 SLISGLFLFLFFWLDORSVLAKELF--VTLVFRHGDRSPIDTPTDPIKSSWPQGFQGL 70  
QY 57 TPRG-----AALVILGYRYF-----ARGLLPGCPAAGTILADVDRTRTGOA 98  
DB 71 TQLGMEQHYELGEYIRKRYKFLNESYKHEQVYIRS-----TDVDRTLMSAMT 118  
QY 99 FAAGLAP--GCAI-----VHA-----GDD-----PIFHGLDTGCLDQ--- 128  
DB 119 NLALFPPEGVSINWPILLMQPIVHTVPLSEDLQLLYPFRNCPRFOLESETLKSEEFQ 178  
QY 129 -----ADAILAAGEGGLTARHGLTLAKVLNFAFASACLECARVVGRLGPIILAS----- 174  
DB 179 KRLHPYKDFIATLKG--LSGLHGQDLFGIWSKVYDPLCYESVHNFTLPSWATEDWTMKLR 236  
QY 175 TLSIFILLEYAOGPMEVWGGRISAEWLLHNAOFLMNRTPYIARATPILIVTALSAPARVL 234  
DB 237 ELSLSLLSLYGIHKQEKSRLOGGVLVNE---ILN---HMKRATQI-----PSYKKLIM 285  
QY 235 LAGHDTNLALGGLDLWQLPQDPETPPGGA-----LVFELMNRVYRVVYQTMQDLRNLEPL 289  
DB 286 YSAHDTTVS--GLQWALDVYNGLLPPYASCHLTLYFEKGEYFVEMY-----RNETQH 337  
QY 290 LPRILPIPGCGSEACSLSDFARLVAP 316  
DB 338 EYPVLMPLGC--SPSCPLERFAELVGP 362

RESULT 11

US-09-097-199-48  
; Sequence 48, Application US/09097199  
; Patent No. 6218529  
; GENERAL INFORMATION:  
; APPLICANT: An, Gang  
; APPLICANT: O'Hara, S. Mark  
; APPLICANT: Veltri, Robert  
; APPLICANT: Veltri, Robert  
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,  
; TITLE OF INVENTION: PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE  
; NUMBER OF SEQUENCES: 87



```
/
/ CORRESPONDENCE ADDRESS:
/ ADDRESSES: Arnold, White & Durkee
/ STREET: P.O. Box 4433
/ CITY: Houston
/ STATE: Texas
/ COUNTRY: USA
/ ZIP: 77210
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent in Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/097,199
/ FILING DATE:
/
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/692,787
/ FILING DATE: 31-JUL-1996
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Nakaehima, Richard A.
/ REGISTRATION NUMBER: P-42,023
/ REFERENCE/DOCKET NUMBER: UROC:018
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (512) 418-3000
/ TELEFAX: (512) 474-7577
/ INFORMATION FOR SEQ ID NO: 48:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 386 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/
/ US-09-097-199-48
/
/ Query Match 6.3%; Score 103.5; DB 3; Length 386;
/ Best Local Similarity 23.8%; Pred. No. 0.0062;
/ Matches 92; Conservative 38; Mismatches 148; Indels 109; Gaps 20;
/
/ QY 2 SLAAALAAALAPAAAAAEPLKLVILSRHGVRSPTKALMVSP--DWFVPVGV---L 56
/ DB 13 SLGLGLFLFLFWLDRSLAKELKF--VTLNFRHGDRSPIDFTPTPIKESWPQGGQL 70
/ QY 57 TPRG-----AALVLGVYP-----AGLLPGCPAAGTILADVDERTRTGOA 98
/ DB 71 TQGMQSHVELGYIKRYKRYKFLNESYKHEQVIRS-----TDVDRTLMSANT 118
/ QY 99 FAAGLAP--GCAT-----VHA---GDD-----PIFHGLDTGCLDQ---- 128
/ DB 119 NLAAALPPGEGSVIWNPIILWQPIVHTVPLSEDDQLYLPFRNCPFOESETLKSEEPQ 178
/ QY 129 -----ADAILAAGEGLTARHGLTLAKVLPASACLECARVVGRLGPLLAS----- 174
/ DB 179 KRLHPYKDFATIGK--LSGLHQDGLFGIWSKYVDLYCESVHNFTLPSWATEDTWTUR 236
/ QY 175 TLSEIFLLEVAQGPMEYGMGRISAEMLLHNAQFLMNRTPYIARATPILIVTALSFARVVL 234
/ DB 237 ELSELSSLISLYGIHKQKESRLOGGVLVNE--ILN---HMKEATQI-----PSYKLLIM 285
/ QY 235 LAGHDNTNALGGDLWLQDPDPTPPGGA-----LVFELNRYVRYVMYQTMGLRLNLEPL 289
/ DB 286 YSAHDTTVS--GLQMALDVYVNGLLPPYASCHLTLYFEKGEYFVEMY-----RNETQH 337
/ QY 290 LPRILPPIPGGSEAAACSLSDFAILVAP 316
/ DB 338 EPIPLMLPGC--SPSCPLERFAELVGP 362
/
/ RESULT 12
/ US-09-252-991A-25845
/ Sequence 25845, Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252,991A
/ CURRENT FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 25845
/ LENGTH: 514
/ TYPE: PRT
/ ORGANISM: Pseudomonas aeruginosa
/
/ US-09-252-991A-25845
/
/ Query Match 6.3%; Score 103.5; DB 4; Length 514;
/ Best Local Similarity 24.7%; Pred. No. 0.0095;
/ Matches 96; Conservative 37; Mismatches 125; Indels 131; Gaps 20;
/
/ QY 1 VSLLLAALLALA-ALAPAAAAAEPLKLVILSRHGVRSPTKALMVSPDW-----PWPV 54
/ DB 86 IALFLPEALFMSLAGGVLSAALPLFLE-----RQGER-----RLDWLAAILPALL 131
/ QY 55 GLTPRGAALVLLGYRYFARGLLPGCPAAGTILADVDERTRTGOAFA-----AGLARGCAI 109
/ DB 132 GIAVALSLLLAAPMLVRLGPG-----LAE-----TASQAQAANLRLVLCVPLG-- 178
/ QY 110 VHAGDDPIFHGLDTGCLDQADAILAAGEGL-----TARHGLTTLAKVLPASACL- 159
/ DB 179 -----MLHALPSIPLQAAERFVLAGLSLLNPLPVLVLAHVG-QASQPEQLALACLL 230
/ QY 160 -----ECARVVGRGLGPELLASTLSE--IFLLEYAQGPM 199
/ DB 231 GSLIMPLVLLPSLWIEGCRPWHRLSGTELGEGLGGRIGPILLNNAASQGLALVERLVASL 290
/ QY 190 EVGGRISAEMLLHNAQFLMNRTPYIARATPILIVTALSFA- 231
/ DB 291 -LGEGAVT--W-VNLARKLMN-LPLIALMSLNQVILGMMRSRQGERLALLRGLTASL 345
/ QY 232 VVLAGHDNTNALGGDLWLQDPDPTPPGGA-----LVFELN-----RYVRYVMYQT 279
/ DB 346 LTLFAGVGLVAAPGLVALLPQQTAGSLPLFALLAWFAVPLVFGVGNALLARYAAGDT 405
/ QY 280 MDQLR-----NLEPLLPRILPIPG 298
/ DB 406 RLPLRCELLGSALNAALLAVLPLIFGLPG 434
/
/ RESULT 13
/ US-09-146-283-2
/ Sequence 2, Application US/09146283
/ Patent No. 5976546
/ GENERAL INFORMATION:
/ APPLICANT: Laus, Reiner
/ APPLICANT: Ruegg, Curtis L.
/ APPLICANT: Wu, Hongyu
/ TITLE OF INVENTION: Immunostimulatory Compositions
/ NUMBER OF SEQUENCES: 10
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Denlinger & Associates
/ STREET: 350 Cambridge Ave, Suite 250
/ CITY: Palo Alto
/ STATE: CA
/ COUNTRY: USA
/ ZIP: 94306
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent in Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/146,283
/
```

```

; FILING DATE: 03-SEPT-1998
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Judge, Linda R.
; REGISTRATION NUMBER: 42,702
; REFERENCE/DOCKET NUMBER: 7636-0010.21
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-0880
; TELEFAX: 650-324-0960
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 515 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; INDIVIDUAL ISOLATE: Prostatic acid phosphatase-GM-CSF
; INDIVIDUAL ISOLATE: fusion protein; Fig. 1
US-09-146-283-2

Query Match 6.3%; Score 103.5; DB 2; Length 515;
Best Local Similarity 23.8%; Pred. No. 0.0096;
Matches 92; Conservative 38; Mismatches 148; Indels 109; Gaps 20;

QY 2 SLLLAALLAALAPAAAAEPLKLVILSRHGVSPKALMVSP--DWPWPVG---L 56
DB 13 SLSGLFLFLFPWLDRLSVLAKELK--VTLVFRHGDRSPIDTPTDPIKESWPGQGL 70
QY 57 TPRG-----AALVLGYRYF-----ARGLLPGCPAAGTILADVDRTRTQQA 98
DB 71 TOLGMEQHYELGEYIRKRYKFLNESYKHEQVIYS-----TDVDRTLMSAMT 118
QY 99 FAAGLAP--GCAI-----VHA---GDD-----PIFHGLDTGCLDQ--- 128
DB 119 NLAAFPPEGVSINWPILLWQIPVHTVPLSEDLQLYLPFRNCPFOESETLKSEEFQ 178
QY 129 -----ADAILAAGEGGITARHGLTLAKVLFASFACLECARVVGRLGPLLAS----- 174
DB 179 KRLHPYKDFIATLKG--LSGLHGQDLFGINSKYVDPLYCESVNFPLPSWATEDTWTCLR 236
QY 175 TLSEIFILEYAGGMEVGNWGRISAEWLLHNAQFLMNRTPYIARATPILIVTALSAPRVL 234
DB 237 ELSELSSLISLYGIHKQKSKRLQGGVLVNE---ILN---HMKRATQI-----PSYKGLIM 285
QY 235 LAGHDTNLALGGLDLWLPQDPETPPGGA-----LVFELWNRVYRVVYQTMQDLRNLEPL 289
DB 286 YSAHDTTVS--GLQMALDVYNGLLPPYASCHLTLYFEKGEYFVEMY-----RNETHQ 337
QY 290 LPRILPIGCGSEAAACSLSDPARLVAP 316
DB 338 EYPPLMLPGC--SPSCPLERFAELVGP 362

RESULT 15
US-09-344-195-2
; Sequence 2, Application US/09344195
; Patent No. 6210662
; GENERAL INFORMATION:
; APPLICANT: Laus, Reiner
; APPLICANT: Wu, Hongyu
; APPLICANT: Ruegg, Curtis L.
; TITLE OF INVENTION: Immunostimulatory Compositions
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; STREET: 350 Cambridge Ave. Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:

```

```

; FILING DATE: 03-SEPT-1998
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Judge, Linda R.
; REGISTRATION NUMBER: 42,702
; REFERENCE/DOCKET NUMBER: 7636-0010.21
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-0880
; TELEFAX: 650-324-0960
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 515 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; INDIVIDUAL ISOLATE: Prostatic acid phosphatase-GM-CSF
; INDIVIDUAL ISOLATE: fusion protein; Fig. 1
US-09-146-283-2

Query Match 6.3%; Score 103.5; DB 2; Length 515;
Best Local Similarity 23.8%; Pred. No. 0.0096;
Matches 92; Conservative 38; Mismatches 148; Indels 109; Gaps 20;

QY 2 SLLLAALLAALAPAAAAEPLKLVILSRHGVSPKALMVSP--DWPWPVG---L 56
DB 13 SLSGLFLFLFPWLDRLSVLAKELK--VTLVFRHGDRSPIDTPTDPIKESWPGQGL 70
QY 57 TPRG-----AALVLGYRYF-----ARGLLPGCPAAGTILADVDRTRTQQA 98
DB 71 TOLGMEQHYELGEYIRKRYKFLNESYKHEQVIYS-----TDVDRTLMSAMT 118
QY 99 FAAGLAP--GCAI-----VHA---GDD-----PIFHGLDTGCLDQ--- 128
DB 119 NLAAFPPEGVSINWPILLWQIPVHTVPLSEDLQLYLPFRNCPFOESETLKSEEFQ 178
QY 129 -----ADAILAAGEGGITARHGLTLAKVLFASFACLECARVVGRLGPLLAS----- 174
DB 179 KRLHPYKDFIATLKG--LSGLHGQDLFGINSKYVDPLYCESVNFPLPSWATEDTWTCLR 236
QY 175 TLSEIFILEYAGGMEVGNWGRISAEWLLHNAQFLMNRTPYIARATPILIVTALSAPRVL 234
DB 237 ELSELSSLISLYGIHKQKSKRLQGGVLVNE---ILN---HMKRATQI-----PSYKGLIM 285
QY 235 LAGHDTNLALGGLDLWLPQDPETPPGGA-----LVFELWNRVYRVVYQTMQDLRNLEPL 289
DB 286 YSAHDTTVS--GLQMALDVYNGLLPPYASCHLTLYFEKGEYFVEMY-----RNETHQ 337
QY 290 LPRILPIGCGSEAAACSLSDPARLVAP 316
DB 338 EYPPLMLPGC--SPSCPLERFAELVGP 362

RESULT 14
US-09-579-823A-2
; Sequence 2, Application US/08579823A
; Patent No. 6080409
; GENERAL INFORMATION:
; APPLICANT: Laus, Reiner
; APPLICANT: Wu, Hongyu
; APPLICANT: Ruegg, Curtis L.
; TITLE OF INVENTION: Immunostimulatory Composition and Method
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; STREET: 350 Cambridge Ave. Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:

```



GenCore version 5.1.1.6  
 Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 12:41:31 ; Search time 31.0244 Seconds  
 (without alignments)  
 2841.193 Million cell updates/sec

Title: US-10-021-723B-14  
 Perfect score: 1647  
 Sequence: 1 VSLLAALLAALAPAAAA.....CGSEAAACSLSDFAFLVAPAC 318

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 1138120 seqs, 277189581 residues

Total number of hits satisfying chosen parameters: 1138120

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US05\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1647	100.0	318	14	US-10-021-723A-14
2	663.5	40.3	409	14	US-10-021-723A-8
3	650	39.5	421	14	US-10-021-723A-6
4	632.5	38.4	430	9	US-09-866-379-10
5	627.5	38.1	436	14	US-10-156-660-2
6	616.5	37.4	412	14	US-10-334-672-1
7	616.5	37.4	412	14	US-10-334-671-1
8	613.5	37.2	410	14	US-10-021-723A-13
9	613.5	37.2	410	14	US-10-021-723A-15
10	613.5	37.2	432	9	US-09-866-379-8
11	613.5	37.2	432	12	US-10-282-122A-43351
12	613.5	37.2	432	14	US-10-156-660-4
13	611.5	37.1	440	9	US-09-777-566A-2
14	611.5	37.1	440	9	US-09-866-379-2
15	611.5	37.1	440	14	US-10-034-985-2

16	611.5	37.1	440	15	US-10-430-356-2	Sequence 2, Appli
17	611	37.1	476	14	US-10-021-723A-10	Sequence 10, Appli
18	608.5	36.9	432	15	US-10-284-962-3	Sequence 3, Appli
19	608.5	36.9	432	15	US-10-284-962-14	Sequence 14, Appli
20	608.5	36.9	433	14	US-10-286-041-1	Sequence 1, Appli
21	602.5	36.6	432	15	US-10-284-962-5	Sequence 5, Appli
22	596	36.2	441	14	US-10-021-723A-12	Sequence 12, Appli
23	590	35.8	441	12	US-10-282-122A-77792	Sequence 77792, A
24	589	35.8	441	14	US-10-021-723A-2	Sequence 2, Appli
25	567	34.4	420	14	US-10-021-723A-4	Sequence 4, Appli
26	554.5	33.7	432	14	US-10-021-723A-16	Sequence 16, Appli
27	138	8.4	426	12	US-10-257-174-44	Sequence 44, Appli
28	138	8.4	426	12	US-10-343-357-6	Sequence 6, Appli
29	117	7.1	502	12	US-10-389-647-684	Sequence 684, App
30	109.5	6.6	3668	15	US-10-402-842-4	Sequence 4, Appli
31	108	6.6	401	15	US-10-369-493-20570	Sequence 20570, A
32	107.5	6.5	801	9	US-09-822-827-982	Sequence 982, App
33	107.5	6.5	801	9	US-09-895-793-982	Sequence 982, App
34	104	6.3	474	16	US-10-354-437-32	Sequence 32, Appli
35	103.5	6.3	386	9	US-09-822-827-945	Sequence 945, App
36	103.5	6.3	386	9	US-09-895-793-945	Sequence 945, App
37	103.5	6.3	386	10	US-09-974-546-48	Sequence 48, Appli
38	103.5	6.3	386	15	US-10-341-434-214	Sequence 214, App
39	102.5	6.2	4551	10	US-09-793-708-1	Sequence 1, Appli
40	102.5	6.2	4551	14	US-10-201-365-2	Sequence 2, Appli
41	102.5	6.2	4551	14	US-10-160-539-1	Sequence 1, Appli
42	102.5	6.2	4613	9	US-09-861-289-31	Sequence 31, Appli
43	102.5	6.2	4613	9	US-09-860-846-31	Sequence 31, Appli
44	102.5	6.2	4613	10	US-09-988-384B-31	Sequence 31, Appli
45	102.5	6.2	4613	10	US-09-836-821-31	Sequence 31, Appli

## ALIGNMENTS

## RESULT 1

US-10-021-723A-14  
 ; Sequence 14, Application US/10021723A  
 ; Publication No. US20030101476A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Short, Jay  
 ; APPLICANT: Mathur, Eric  
 ; APPLICANT: Richardson, Toby  
 ; APPLICANT: Robertson, Dan  
 ; APPLICANT: Barton, Nelson  
 ; TITLE OF INVENTION: Recombinant Phytases and Uses Thereof  
 ; FILE REFERENCE: 112766.140 (DIV.016C1P)  
 ; CURRENT APPLICATION NUMBER: US/10/021,723A  
 ; CURRENT FILING DATE: 2002-10-23  
 ; PRIOR APPLICATION NUMBER: US 60/255,090  
 ; PRIOR FILING DATE: 2000-12-12  
 ; NUMBER OF SEQ ID NOS: 74  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 14  
 ; LENGTH: 318  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: consensus sequence  
 US-10-021-723A-14

Query Match 100.0%; Score 1647; DB 14; Length 318;  
 Best Local Similarity 100.0%; Pred. No. 66-149;  
 Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSLLAALLAALAPAAAAAEPKLEKVVILSRHGVSRPTKALMVSPDPWPNVGLTPRG 60

DB 1 VSLLAALLAALAPAAAAAEPKLEKVVILSRHGVSRPTKALMVSPDPWPNVGLTPRG 60

QY 61 AALVLLGYRFARGLLPGCPAAGTILADVDERITGTQAFAGLAPGCAIVHAGDDPIFHG 120

DB 61 AALVLLGYRFARGLLPGCPAAGTILADVDERITGTQAFAGLAPGCAIVHAGDDPIFHG 120

QY 121 LDTGCLDQADAILAAGEGLTARHGLTLAKVNFASACLECARVVGRLGPILASTLSEIF 180  
Db 121 LDTGCLDQADAILAAGEGLTARHGLTLAKVNFASACLECARVVGRLGPILASTLSEIF 180  
QY 181 LLEYAQGPMEVWGWRISAEWLLHNAQFLMNRTPYIARATPILIVTALSAPARVLLAGHDT 240  
Db 181 LLEYAQGPMEVWGWRISAEWLLHNAQFLMNRTPYIARATPILIVTALSAPARVLLAGHDT 240  
QY 241 NLALGGLDLWLPQDETPPGGALVFELWNRVVRVMTDQRLNLEPLLRIPILPGCG 300  
Db 241 NLALGGLDLWLPQDETPPGGALVFELWNRVVRVMTDQRLNLEPLLRIPILPGCG 300  
QY 301 SEACSLSDFAELVAPAC 318  
Db 301 SEACSLSDFAELVAPAC 318  
RESULT 2  
US-10-021-723A-8  
; Sequence 8, Application US/10021723A  
; Publication No. US20030101476A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay  
; APPLICANT: Mathur, Eric  
; APPLICANT: Richardson, Toby  
; APPLICANT: Robertson, Dan  
; APPLICANT: Barton, Nelson  
; TITLE OF INVENTION: Recombinant Phytases and Uses Thereof  
; FILE REFERENCE: 112766.140 (DIV-016CIP)  
; CURRENT APPLICATION NUMBER: US/10/021,723A  
; CURRENT FILING DATE: 2002-10-23  
; PRIOR APPLICATION NUMBER: US 60/255,090  
; PRIOR FILING DATE: 2000-12-12  
; NUMBER OF SEQ ID NOS: 74  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 409  
; TYPE: PRT  
; ORGANISM: Rhizobium  
US-10-021-723A-8  
Query Match 40.3%; Score 663.5; DB 14; Length 409;  
Best Local Similarity 54.0%; Pred. No. 8.8e-55;  
Matches 218; Conservative 22; Mismatches 75; Indels 89; Gaps 39;  
QY 2 SILL-AALLA-LAALAPAA-----AAEPLKLEKVVILSRHGVRSPTKALMVSP-----DW 50  
Db 4 SILLPALLAGCAASAPACASPAAPGSLKLEKVVMLRHGVRSPTKAAVVPVPGYSATW 63  
QY 51 P-WPYG---LTPRGAALV-LIGYR---YF-ARGLLP-GCPAAGTIL--ADVDERT-RTGQ 97  
Db 64 PDWPVDFGLLTPHGAAGVKLLGESDRLYFGGRGLFPFGCPAAGTIVLKASYKERTIATAQ 123  
QY 98 AFAAGLAGCCA--IVH-AG--DDPIFHGLDTGCILD-----QADAILA-AGEGGLTA---- 142  
Db 124 SWAAGFMGCTADVAHPAGPDDDAIFHGLDGGPASFDGKRAFDAAALQAPEGGLTASTAR 183  
QY 143 -RHGLT-LAKVNFSA-SAC---LECARVVGSR-----LGPL-LASTLSIFLLEYAQG- 187  
Db 184 HRGELTLAKVNLNCPALPACPIAEPRLVAQPHDRPELEGDLVGSTASQTLVLEYLEGK 243  
QY 188 PM-EVWGGRIS---AEWLL--HNAQF-LMNRTPYI--RATPIL--IVTAL---SPARVV 233  
Db 244 PMAEVWGGRVSRABTEQLLRHPLKFRYSNRPYIAATAAAPIVREIVTALGDRSPARLT 303  
QY 234 LLAGHDTNLA-LGG-LDL-WOLPO--PDETPPGGALVFEL-----WNRVVRVMT--OTMD 281  
Db 304 LLAGHDTNVDLGGFFDLHWQVPSYPADEVPPGSGALGFELVSNKAGDRYVRAFIRGQTM 363  
QY 282 QLRNLEP-----LLPRILPIPGCGSE---AACSLSDFAELVAP 316  
Db 364 QLRNLEPRLSDTLFRYRLPIPGCGNSVEATACAWSDFARLAAP 407

## RESULT 3

US-10-021-723A-6  
; Sequence 6, Application US/10021723A  
; Publication No. US20030101476A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay  
; APPLICANT: Mathur, Eric  
; APPLICANT: Richardson, Toby  
; APPLICANT: Robertson, Dan  
; APPLICANT: Barton, Nelson  
; TITLE OF INVENTION: Recombinant Phytases and Uses Thereof  
; FILE REFERENCE: 112766.140 (DIV-016CIP)  
; CURRENT APPLICATION NUMBER: US/10/021,723A  
; CURRENT FILING DATE: 2002-10-23  
; PRIOR APPLICATION NUMBER: US 60/255,090  
; PRIOR FILING DATE: 2000-12-12  
; NUMBER OF SEQ ID NOS: 74  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 421  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: 953-6 phytase sequence  
US-10-021-723A-6

Query Match 39.5%; Score 650; DB 14; Length 421;  
Best Local Similarity 53.0%; Pred. No. 1.8e-53;  
Matches 213; Conservative 20; Mismatches 81; Indels 88; Gaps 37;  
QY 3 LLLAALLA-LAALAPAA-----AAEPLKLEKVVILSRHGVRSPTKALMVSP-----DWP- 51  
Db 18 LLSAALLAGCAASAPACASPAAPGSLKLEKVVMLRHGVRSPTKAAVVPVPGYSATWPD 77  
QY 52 WPVG---LTPRGAALV-LIG---YRFARGLLP-GCPAAGTIL--ADVDERT-RTGQAF 99  
Db 78 WPVDFGLLTPHGAAGVKLLGESDRLTFGGRGLFPDGCFAAGTIVLKASYKERTIATAQNW 137  
QY 100 AAGLAGCCA--IVH-AG--DDPIFHGLDTGCILD-----QADAILA-AGEGGLTA----R 143  
Db 138 AAGFMGCTADVAHPAGPDDDAIFHGLDGGPASFDGKRAFDAAALQAPEGGLTASTARHR 197  
QY 144 HGLT-LAKVNFSA-SAC---LECARVVGSR-----LGPL-LASTLSIFLLEYAQG-PM 189  
Db 198 GELTLAKVNLNCPALPACPIAEPRLVAQPHDRPELEGDLVGSTASQTLVLEYLEGKPM 257  
QY 190 -EYWGGRIS---AEWLL--HNAQF-LMNRTPYI--ARATPIL--IVTAL---SPARVLL 235  
Db 258 AEVWGGRVSRABTEQLLRHPLKFRYSNRPYIAATAAAPIVREIVTALGDRSPARLTIL 317  
QY 236 AGHDTNLA-LGG-LDL-WOLPO--PDETPPGGALVFEL-----WNRVVRVMT--OTMDOL 283  
Db 318 AGHDTNVDLGGFFDLHWQVPSYPADEVPPGSGALGFELVSNKAGDRYVRAFIRGQTM 377  
QY 284 RNLEP-----LLPRILPIPGCG---SEACSLSDFAELVAP 316  
Db 378 RNLEPLGSDALYRRLPIPGCGHSVEATACAWSDFARLAAP 419

## RESULT 4

US-09-866-379-10  
; Sequence 10, Application US/09866379  
; Patent No. US20020136754A1  
; GENERAL INFORMATION:  
; APPLICANT: DIVERSA CORPORATION  
; APPLICANT: Short, Jay  
; APPLICANT: KRETZ, Keith  
; APPLICANT: GRAY, Kevin  
; APPLICANT: BARTON, Nelson  
; APPLICANT: GARRETT, James  
; APPLICANT: O'DONOGHUE, Eileen  
; TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYTASES AND USES THEREOF

```
; FILE REFERENCE: DIVER1370-7
; CURRENT APPLICATION NUMBER: US/09/866,379
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 09/580,515
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: US 08/259,214
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: US 08/910,798
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Modified phytase
US-09-866-379-10

Query Match      38.4%; Score 632.5; DB 9; Length 430;
Best Local Similarity 45.2%; Pred. No. 8.6e-52;
Matches 194; Conservative 44; Mismatches 76; Indels 115; Gaps 35;

QY  2 SLLIAALLAALAP--AAAAAEP-LKLEKVVILSRHGVSRPTKA--LM--VSPD-WP-W 52
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  3 AILIPFLSLILPTPQSAFAQSEPELKLESVIVSRHGVRAFTKATQLMQDVTPDWPTW 62

QY  53 PV---GLTPRGAALV-LLGY-----RYFARGLLP--GCPAAG--TILADVDERTR-TGOAF 99
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  63 PVKLGELTPRGELIAYLGHYWRQLVADGLLPKCGCPQSGQVAILADVDERTRKTGEAF 122

QY  100 AAGLAPGCAI-VH-----AGDDPIFHGLDTG-C-LDOA---DAILAAGEGLTARHG--- 145
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  123 AAGLAPCAITVHTQADTSSPDPLFNPLKTVGCOLDNANVTDAILEAG-GSIADFTGHYQ 181

QY  146 ---LTLAKVLNF--ASACL-----ECARVVGSLGPELLASTLS 177
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  182 TAFRELERVLNFPQSNCLKREKQDECSLTQALPSELKVSADCVSLTGAVS--LASMLT 239

QY  178 EIFLELYAOGPMVEGWGRI--SAEW-----LLHNAQFLMNETPYIA--RATPI--LIVTA 227
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  240 EIFLQQAQMPGPGWGRITDSHQWNTLLSLHNAQFOLLQRTPEVARSRATPLDLIKTA 299

QY  228 S-----PARVVLLAGHDTNLA-LGG-LDL-WOLP-QPDETTPGGALVFELWN 270
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  300 TEPHPQQAQYGVTLFTSVLFIAGHDTNLANLGGALELNWTLPGQPDNTPGGELVFERWR 359

QY  271 R-----YRVVMYQTMQDLRLNLEPLLPRILP-----IPGC---GSEAACSLSDFARL 313
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  360 RLSDNSQWIQVSLVFQTLQQRDKTPLUSLNTPPGEVKLTLAGCEERNAQMGCSLAGFTQI 419

QY  314 V----APAC 318
   : : : : :
Db  420 VNEARIAPAC 428

RESULT 5
US-10-156-660-2
; Sequence 2, Application US/10156660
; Publication No. US20030103958A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay M.
; APPLICANT: Kretz, Keith
; APPLICANT: Gray, Kevin A.
; APPLICANT: Barton, Nelson R.
; APPLICANT: Garrett, James B.
; APPLICANT: O'Donoghue, Bileen
; APPLICANT: Mathur, Eric J.
; TITLE OF INVENTION: PHYTASES, NUCLEIC ACIDS ENCODING THEM
```

```
; TITLE OF INVENTION: AND METHODS FOR MAKING AND USING THEM
; FILE REFERENCE: 09010-029007
; CURRENT APPLICATION NUMBER: US/10/156,660
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: US 09/866,379
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 09/580,515
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: US 09/259,214
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: US 08/910,798
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 436
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: modified phytase enzyme
US-10-156-660-2

Query Match      38.1%; Score 627.5; DB 14; Length 436;
Best Local Similarity 44.9%; Pred. No. 2.6e-51;
Matches 193; Conservative 44; Mismatches 78; Indels 115; Gaps 35;

QY  2 SLLIAALLAALAP--AAAAAEP-LKLEKVVILSRHGVSRPTKA--LM--VSPD-WP-W 52
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  3 AILIPFLSLILPTPQSAFAQSEPELKLESVIVSRHGVRAFTKATQLMQDVTPDWPTW 62

QY  53 PV---GLTPRGAALV-LLGY-----RYFARGLLP--GCPAAG--TILADVDERTR-TGOAF 99
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  63 PVKLGELTPRGELIAYLGHYWRQLVADGLLPKCGCPQSGQVAILADVDERTRKTGEAF 122

QY  100 AAGLAPGCAI-VH-----AGDDPIFHGLDTG-C-LDOA---DAILAAGEGLTARHG--- 145
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  123 AAGLAPCAITVHTQADTSSPDPLFNPLKTVGCOLDNANVTDAILEAGGSIADFTGHYQ 182

QY  146 ---LTLAKVLNF--ASACL-----ECARVVGSLGPELLASTLS 177
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  183 TAFRELERVLNFPQSNCLKREKQDECSLTQALPSELKVSADCVSLTGAVS--LASMLT 240

QY  178 EIFLELYAOGPMVEGWGRI--SAEW-----LLHNAQFLMNETPYIA--RATPI--LIVTA 226
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  241 EIFLQQAQMPGPGWGRITDSHQWNTLLSLHNAQFOLLQRTPEVARSRATPLDLIKTA 300

QY  227 LS-----PARVVLLAGHDTNLA-LGG-LDL-WOLP-QPDETTPGGALVFELW 269
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  301 LTPHPQQAQYGVTLFTSVLFIAGHDTNLANLGGALELNWTLPGQPDNTPGGELVFERW 360

QY  270 NR-----YRVVMYQTMQDLRLNLEPLLPRILP-----IPGC---GSEAACSLSDFAR 312
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  361 RLSDNSQWIQVSLVFQTLQQRDKTPLUSLNTPPGEVKLTLAGCEERNAQMGCSLAGFTQ 420

QY  313 LV----APAC 318
   : : : : :
Db  421 IVNEARIAPAC 430

RESULT 6
US-10-334-672-1
; Sequence 1, Application US/10334672
; Publication No. US20030157646A1
; GENERAL INFORMATION:
; APPLICANT: Lanahan, Mike
; APPLICANT: Koepf, Edward
; APPLICANT: Kretz, Keith
; TITLE OF INVENTION: Microbially-Expressed Thermotolerant Phytase For Animal Feed
; FILE REFERENCE: SYNG-P01-001
```

```
/ CURRENT APPLICATION NUMBER: US/10/334,672
/ CURRENT FILING DATE: 2002-12-30
/ PRIOR APPLICATION NUMBER: 60/344,523
/ PRIOR FILING DATE: 2001-12-28
/ NUMBER OF SEQ ID NOS: 4
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 1
/ LENGTH: 412
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: No. US20030157646A19X Phytase
US-10-334-672-1

Query Match      37.4%; Score 616.5; DB 14; Length 412;
Best Local Similarity 45.5%; Pred. No. 2.8e-50;
Matches 187; Conservative 41; Mismatches 70; Indels 113; Gaps 34;

QY 19 AAAP-LKLEKVVILSRHGVRSPTKA--LM--VSPD-WP-WPV---GLTPRGAALV-LIG 67
Db 2 AQSEPELKLESVIVSRHGVRAPTKATQMQDVTDPAMTPVVKLGELTPRGGLIAYLG 61

QY 68 Y----RYFARGLLP--GCPAAG--TILADVDBTR--TQAFAGAGLAPGCAI-VH-----A 112
Db 62 HYWRQLVADGLLPKCGCPQSGQVAIIADVDBTRTKTGEAFAGLAPDCAITVHTQADTS 121

QY 113 GDDPIFHGLDTG-C-LDQA---DAILAGEGGLTARHG-----LTLAKVLNF--ASACL 159
Db 122 SPDPLFNPLKTVGCOLDNANVTDAILERAGGSIADFTGHYQTAFARELERVLNFPQSNLCL 181

QY 160 -----ECARVVGRIGLPLLASTLSEIFLLEYAAGPMVEVWGRI 196
Db 182 KREKQDESCSLTQALPSELKVSADCVSLTGAVS--LASMLTEIFLQQAQGMPEFGWRI 239

QY 197 --SAEW---LHNAQF-LMNRTPYIA--RATPI--LIVTALS-----PARV 232
Db 240 TDSHOMNTLLSHNAQFDLLQRTPEVARSRAIPLDLIKTALTTPHPQQAQVGLTPTSV 299

QY 233 VLLAGHDTNLA-LGG-LDL-WQLP-QPDETPPGGALVFELMNR-----YVRVMYQTM 280
Db 300 LFIAGHDTNLANLGGALELNWTLPGQDNTPPGGELVFERWRLSDNSQWIVQSLVFQTL 359

QY 281 DQIRNLEPLLPRILP-----IPGC---GSEAAACSLSDPARLV-----APAC 318
Db 360 QQMRDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIIPAC 410

RESULT 7
US-10-334-671-1
/ Sequence 1, Application US/10334671
/ Publication No. US20030170293A1
/ GENERAL INFORMATION:
/ APPLICANT: Lanahan, Mike
/ APPLICANT: Koepf, Edward
/ APPLICANT: Kretz, Keith
/ TITLE OF INVENTION: Microbially-Expressed Thermotolerant Phytase For Animal Feed
/ FILE REFERENCE: 70098
/ CURRENT APPLICATION NUMBER: US/10/334,671
/ CURRENT FILING DATE: 2002-12-30
/ PRIOR APPLICATION NUMBER: 60/344,523
/ PRIOR FILING DATE: 2001-12-28
/ NUMBER OF SEQ ID NOS: 4
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 1
/ LENGTH: 412
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: No. US20030170293A19X Phytase
US-10-334-671-1

Query Match      37.4%; Score 616.5; DB 14; Length 412;
Best Local Similarity 45.5%; Pred. No. 2.8e-50;
```

```
Matches 187; Conservative 41; Mismatches 70; Indels 113; Gaps 34;

QY 19 AAAP-LKLEKVVILSRHGVRSPTKA--LM--VSPD-WP-WPV---GLTPRGAALV-LIG 67
Db 2 AQSEPELKLESVIVSRHGVRAPTKATQMQDVTDPAMTPVVKLGELTPRGGLIAYLG 61

QY 68 Y----RYFARGLLP--GCPAAG--TILADVDBTR--TQAFAGAGLAPGCAI-VH-----A 112
Db 62 HYWRQLVADGLLPKCGCPQSGQVAIIADVDBTRTKTGEAFAGLAPDCAITVHTQADTS 121

QY 113 GDDPIFHGLDTG-C-LDQA---DAILAGEGGLTARHG-----LTLAKVLNF--ASACL 159
Db 122 SPDPLFNPLKTVGCOLDNANVTDAILERAGGSIADFTGHYQTAFARELERVLNFPQSNLCL 181

QY 160 -----ECARVVGRIGLPLLASTLSEIFLLEYAAGPMVEVWGRI 196
Db 182 KREKQDESCSLTQALPSELKVSADCVSLTGAVS--LASMLTEIFLQQAQGMPEFGWRI 239

QY 197 --SAEW---LHNAQF-LMNRTPYIA--RATPI--LIVTALS-----PARV 232
Db 240 TDSHOMNTLLSHNAQFDLLQRTPEVARSRAIPLDLIKTALTTPHPQQAQVGLTPTSV 299

QY 233 VLLAGHDTNLA-LGG-LDL-WQLP-QPDETPPGGALVFELMNR-----YVRVMYQTM 280
Db 300 LFIAGHDTNLANLGGALELNWTLPGQDNTPPGGELVFERWRLSDNSQWIVQSLVFQTL 359

QY 281 DQIRNLEPLLPRILP-----IPGC---GSEAAACSLSDPARLV-----APAC 318
Db 360 QQMRDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIIPAC 410

RESULT 8
US-10-021-723A-13
/ Sequence 13, Application US/10021723A
/ Publication No. US20030101476A1
/ GENERAL INFORMATION:
/ APPLICANT: Short, Jay
/ APPLICANT: Richardson, Toby
/ APPLICANT: Robertson, Dan
/ APPLICANT: Barton, Nelson
/ TITLE OF INVENTION: Recombinant Phytases and Uses Thereof
/ FILE REFERENCE: 112766.140 (DIV-016CIP)
/ CURRENT APPLICATION NUMBER: US/10/021,723A
/ CURRENT FILING DATE: 2002-10-23
/ PRIOR APPLICATION NUMBER: US 60/255,090
/ NUMBER OF SEQ ID NOS: 74
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 13
/ LENGTH: 410
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: amino acid sequence of a phytase enzyme with
/ OTHER INFORMATION: predicted glycosylation sites
US-10-021-723A-13

Query Match      37.2%; Score 613.5; DB 14; Length 410;
Best Local Similarity 45.5%; Pred. No. 5.3e-50;
Matches 186; Conservative 41; Mismatches 69; Indels 113; Gaps 34;

QY 21 AEP-LKLEKVVILSRHGVRSPTKA--LM--VSPD-WP-WPV---GLTPRGAALV-LIGY- 68
Db 2 SEPELKLESVIVSRHGVRAPTKATQMQDVTDPAMTPVVKLGELTPRGGLIAYLGHY 61

QY 69 ---RYFARGLLP--GCPAAG--TILADVDBTR--TQAFAGAGLAPGCAI-VH-----AGD 114
Db 62 WRQRLVADGLLPKCGCPQSGQVAIIADVDBTRTKTGEAFAGLAPDCAITVHTQADTSSP 121

QY 115 DPFPHGLDTG-C-LDQA---DAILAGEGGLTARHG-----LTLAKVLNF--ASACL-- 159
Db 122 DPLFNPLKTVGCOLDNANVTDAILERAGGSIADFTGHYQTAFARELERVLNFPQSNLCL 181
```





```
Db 423 NEARIPAC 430

RESULT 11
US-10-282-122A-43351
; Sequence 43351, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELIURA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43351
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-282-122A-43351

Query Match 37.2%; Score 613.5; DB 12; Length 432;
Best Local Similarity 45.3%; Pred. No. 5.7e-50;
Matches 194; Conservative 44; Mismatches 79; Indels 111; Gaps 36;

QY 2 SLILAAALALAAAP--AAAAAP-LKLEKVVILSRHGVSPTKA--LM--VSPD-WP-W 52
Db 3 AILIPFSLILPITPQSAFAQSEPELKESVIVSRHGVRAPTKATQMDVTPDANPTW 62
QY 53 PV---GLTPRGAALV-LLGY-----RYFARGLL--PGCPAAG--TILADVDERTR-TGOAF 99
Db 63 PVKLGWLTTPRGGLIAYLGHYQRLVADGLLAKKGCPSQGVAVIADVDERTRKTGEAF 122
QY 100 AAGLAPGCAT-VH-----AGDDPIFHGLDTG-C-LDOA---DAILAAGEGGLT--ARHGL 146
Db 123 AAGLAPDCATVHTQADTSSPDLFNLKTVGVCLDNANVTDAILSRAGSADFTGHRQ 182
QY 147 T-----LAKVLNF--ASACLE-----CARVVGRLGFL--LASTLSEI 179
Db 183 TAFRELERVLNFPQSNLCKREKQDESCSLTQALPSKLVSDNVSITGAVSLASMLTEI 242

RESULT 12
US-10-156-660-4
; Sequence 4, Application US/10156660
; Publication No. US20030103958A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay M.
; APPLICANT: Kretz, Keith
; APPLICANT: Gray, Kevin A.
; APPLICANT: Barton, Nelson R.
; APPLICANT: Garrett, James B.
; APPLICANT: O'Donoghue, Eileen
; APPLICANT: Mathur, Eric J.
; TITLE OF INVENTION: PHYTASES, NUCLEIC ACIDS ENCODING THEM
; TITLE OF INVENTION: AND METHODS FOR MAKING AND USING THEM
; FILE REFERENCE: 09010-029007
; CURRENT APPLICATION NUMBER: US/10/156,660
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: US 09/866,379
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 09/580,515
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: US 09/259,214
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: US 08/910,798
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-156-660-4

Query Match 37.2%; Score 613.5; DB 14; Length 432;
Best Local Similarity 45.3%; Pred. No. 5.7e-50;
Matches 194; Conservative 44; Mismatches 79; Indels 111; Gaps 36;

QY 2 SLILAAALALAAAP--AAAAAP-LKLEKVVILSRHGVSPTKA--LM--VSPD-WP-W 52
Db 3 AILIPFSLILPITPQSAFAQSEPELKESVIVSRHGVRAPTKATQMDVTPDANPTW 62
QY 53 PV---GLTPRGAALV-LLGY-----RYFARGLL--PGCPAAG--TILADVDERTR-TGOAF 99
Db 63 PVKLGWLTTPRGGLIAYLGHYQRLVADGLLAKKGCPSQGVAVIADVDERTRKTGEAF 122
QY 100 AAGLAPGCAT-VH-----AGDDPIFHGLDTG-C-LDOA---DAILAAGEGGLT--ARHGL 146
Db 123 AAGLAPDCATVHTQADTSSPDLFNLKTVGVCLDNANVTDAILSRAGSADFTGHRQ 182
QY 147 T-----LAKVLNF--ASACLE-----CARVVGRLGFL--LASTLSEI 179
Db 183 TAFRELERVLNFPQSNLCKREKQDESCSLTQALPSKLVSDNVSITGAVSLASMLTEI 242
```



Db 303 PHPPQKAYGVLTFTSVLFIAGHTDNLNIGALELNTLPGQPDNTPPGGELVFERWR 362  
QY 272 -----YRVVYQTMQDLRLNLEPLPRILP-----IPGC---GSEAAACSLSDPARLV 314  
Db 363 LSDNSQWIOVSLVFQTLQWRDKTPLSLNTPGCEVVKLTLAGCEERNAQGMCSLAGFTQIV 422  
QY 315 -----APAC 318  
Db 423 NEARIPAC 430

## RESULT 15

US-10-034-985-2  
; Sequence 2, Application US/10034985  
; Publication No. US20030049815A1  
; GENERAL INFORMATION:  
; APPLICANT: Kretz, Keith  
; TITLE OF INVENTION: NOVEL PHYTASE  
; FILE REFERENCE: 09010/029003  
; CURRENT APPLICATION NUMBER: US/10/034,985  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: US/09/580,515  
; PRIOR FILING DATE: 1999-05-25  
; PRIOR APPLICATION NUMBER: 09/291,931  
; PRIOR FILING DATE: 1999-04-13  
; PRIOR APPLICATION NUMBER: 08/910,798  
; PRIOR FILING DATE: 1997-08-13  
; PRIOR APPLICATION NUMBER: 08/259,214  
; PRIOR FILING DATE: 1999-03-01  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 440  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-10-034-985-2

Query Match 37.1%; Score 611.5; DB 14; Length 440;  
Best Local Similarity 45.1%; Pred. No. 9e-50;  
Matches 193; Conservative 45; Mismatches 79; Indels 111; Gaps 36;

QY 2 SLLAALLAALAP--AAAAEP-LKLEKVTLSRHGVRSPTKA--LM--VSPD-WP-W 52  
Db 3 AILIPFLSLIPIPTPOSAPQAQSEPELKSIVSRHGVRAPTKATQLMQDVTDPADPTW 62  
QY 53 PV---GLTPGAALV-LLGY---RYFARGLL--PGCPAAG--TILADVDERTR-TGOAF 99  
Db 63 FVKLGWLTFRGGELIAYLGHYQRLVADGLLAKGCPQSGQVAILADVDERTKRGAF 122  
QY 100 AAGLAPGCAI-VH-----AGDDPIFHGLDTG-C-IDQA---DAILAAGEGGLT--ARHGL 146  
Db 123 AAGLAPDCAITVHTQADTSSPDPLFNPLKTVGCQLDNANVTDAILSRAGGSIADFTHRQ 182  
QY 147 T-----LAKVLNF--ASACLE-----CARVVGRLGFL--LASTLSI 179  
Db 183 TAFRELRLVNFQSNCLKREKQDSCSLTQALPSLKVSDNVSLTGAVSLASMLTEI 242  
QY 180 FLLEYAQGPNEVGWRI--SAEW-----LLHNAQF-LMNRTPYIA--RATPI--LIVTALS 228  
Db 243 FLIQQAQGMPEPGMGRITDSHOWNTLSLHNAQFYLLQRTPEVARSRATPLDLIMAALT 302  
QY 229 -----PARVVLGAGHTNLA-LGG-LDL-WOLP-QPDETTPGGALVPELMNR 271  
Db 303 PHPPQKAYGVLTFTSVLFIAGHTDNLNIGALELNTLPGQPDNTPPGGELVFERWR 362  
QY 272 -----YRVVYQTMQDLRLNLEPLPRILP-----IPGC---GSEAAACSLSDPARLV 314  
Db 363 LSDNSQWIOVSLVFQTLQWRDKTPLSLNTPGCEVVKLTLAGCEERNAQGMCSLAGFTQIV 422  
QY 315 -----APAC 318  
Db 423 NEARIPAC 430

Search completed: April 30, 2004, 12:55:05  
Job time : 33.0244 secs

GenCore version 5.1.6  
Copyright (c) 1995 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 12:35:05 ; Search time 11.6341 Seconds  
(without alignments)  
2629.235 Million cell updates/sec

Title: US-10-021-723B-14

Perfect score: 1647

Sequence: 1 VSLIALALLAALAPAAAA.....CGSEAAACSLDFAFLVAPAC 318

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Pir1.\*

2: Pir2.\*

3: Pir3.\*

4: Pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	613.5	37.2	432	2 B36733	acid phosphatase (
2	606.5	36.8	434	2 H90770	phosphoanhydride p
3	606.5	36.8	444	2 D85633	hypothetical prote
4	590	35.8	441	2 AC0201	acid phosphatase (
5	377.5	22.9	414	2 E87316	periplasmic phosph
6	289.5	17.6	413	2 F90773	periplasmic glucos
7	289.5	17.6	413	2 B95636	periplasmic glucos
8	288.5	17.5	413	2 JV0087	Glucose-1-phosphat
9	278.5	16.9	413	2 AC0632	Glucose-1-phosphat
10	256	15.5	417	2 S25627	Glucose-1-phosphat
11	134.5	8.2	423	2 A33395	acid phosphatase (
12	127.5	7.7	421	2 S14742	acid phosphatase (
13	124.5	7.6	423	1 S06167	acid phosphatase (
14	117	7.1	502	2 B93353	xylofuranase PA
15	113.5	6.9	657	2 H70872	probable ctpd prot
16	107.5	6.5	426	2 E87385	chromate transport
17	106	6.4	972	2 E97421	NADH dehydrogenase
18	106	6.4	994	2 AD2639	pH adaption potass
19	104	6.3	358	2 D72637	hypothetical prote
20	103.5	6.3	386	1 JH0610	acid phosphatase (
21	102.5	6.2	4613	2 T17409	polyketide synthas
22	102	6.2	356	2 E84199	hypothetical prote
23	101.5	6.2	532	2 E87343	ABC transporter, A
24	99	6.0	426	2 E83172	probable transport
25	98.5	6.0	411	2 S41945	hypothetical prote
26	98	6.0	402	1 S25077	monensin polyketid
27	98	6.0	409	2 B70636	probable fadE17 pr
28	98	6.0	419	2 H70863	hypothetical prote
29	96.5	5.9	438	2 S64682	acid phosphatase (

30 96 5.8 478 2 C72658 hypothetical prote  
31 95.5 5.8 761 1 S22769 subtilisin-like pr  
32 95 5.8 358 2 T36783 probable zinc-bind  
33 95 5.8 832 2 B70507 hypothetical prote  
34 93.5 5.7 469 2 B83368 hypothetical prote  
35 93.5 5.7 571 2 C84169 molybdenum cofacto  
36 93.5 5.7 693 2 C87575 sensor histidine k  
37 93.5 5.7 1182 2 T29097 pro-pol-dUTPase po  
38 92.5 5.6 893 2 E87244 conserved hypoteth  
39 92.5 5.6 1398 2 T18350 probable pol polyp  
40 92 5.6 288 2 AH1171 transcription regu  
41 92 5.6 560 2 G87613 pilus assembly pro  
42 92 5.6 1098 2 G70697 probable arabinosy  
43 92 5.6 1165 2 D59433 C. elegans protein  
44 91.5 5.6 245 2 E87273 tRNA (Guanine-N1)-  
45 91.5 5.6 520 2 T36454 probable oxidoredu

#### ALIGNMENTS

##### RESULT 1

B36733  
acid phosphatase (EC 3.1.3.2) precursor - Escherichia coli (strain K-12)  
N;Alternate names: phosphoanhydride phosphohydrolase, periplasmic; phytase 6  
C;Species: Escherichia coli  
C;Date: 19-Apr-1991 #sequence revision 19-Apr-1991 #text change 01-Mar-2002  
C;Accession: B36733; S18018; B64839; A26534; S17960; S33278  
R;Dassa, J.; Marck, C.; Boquet, P.L.  
J;Bacteriol. 172, 5497-5500, 1990  
A;Title: The complete nucleotide sequence of the Escherichia coli gene appA reveals sh  
A;Reference number: A36733; MUID:90368616; PMID:2168385  
A;Accession: B36733  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-432 <DAS>  
A;Cross-references: GB:M58708; NID:g145283; PIDN:AAA72086.1; PID:g145285  
R;Greiner, R.; Jany, K.D.  
Biol. Chem. Hoppe-Seyler 372, 664-665, 1991  
A;Title: Characterization of a phytase from Escherichia coli.  
A;Reference number: S18018  
A;Accession: S18018  
A;Molecule type: protein  
A;Residues: 23-33 <GRE>  
R;Blatner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;  
.A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A;Title: The complete genome sequence of Escherichia coli K-12.  
A;Reference number: A64720; MUID:97426617; PMID:9278503  
A;Accession: B64839  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-432 <BLAT>  
A;Cross-references: GB:AF000200; GB:U00096; NID:92367111; PIDN:AAC74065.1; PID:g178721  
A;Experimental source: strain K-12, substrain MG1655  
R;Touati, E.; Danchin, A.  
Biochimie 69, 215-221, 1987  
A;Title: The structure of the promoter and amino terminal region of the pH 2.5 acid ph  
A;Reference number: A26534; MUID:87271766; PMID:3038201  
A;Accession: A26534  
A;Molecule type: DNA  
A;Residues: 1-50, 'NAGCHPRMANIAG', 65, 'T', 67-74, 'DV', 77-111, 'S', <TOU>  
A;Cross-references: GB:X05471; NID:g40925; PIDN:CAA29031.1; PID:g40927  
R;Dassa, J.; Fathi, H.; Marck, C.; Dion, M.; Kieffer-Bontemps, M.; Boquet, P.L.  
Mol. Gen. Genet. 229, 341-352, 1991  
A;Title: A new oxygen-regulated operon in Escherichia coli comprises the genes for a p  
A;Reference number: S17958; MUID:92049231; PMID:1658595  
A;Accession: S17960  
A;Status: not compared with conceptual translation  
A;Molecule type: DNA  
A;Residues: 1-17 <DA2>  
A;Cross-references: GB:S63811; NID:g238656; PIDN:AAB20286.1; PID:g238659  
R;Greiner, R.; Konietzny, U.; Jany, K.D.

Arch. Biochem. Biophys. 303, 107-113, 1993  
A;Title: Purification and characterization of two phytases from *Escherichia coli*.  
A;Reference number: S33278; MUID:93256556; PMID:8387749  
A;Accession: S33278  
A;Molecule type: protein  
A;Residues: 23-31; A, 33-34 <R2>  
C;Comment: In addition to cAMP-mediated control, this enzyme is induced when bacterial  
nasearobic conditions.  
C;Genetics:

A;Gene: appA  
A;Map position: 25 min  
A;Keywords: monomer; periplasmic space; phosphatidyl; phosphoprotein; phosphoric mon  
F;1-22/Domain: signal sequence #status predicted <SIG>  
F;23-112/Product: acid phosphatase #status predicted <MAT>  
F;38/Active site: Arg #status predicted  
F;39/Active site: His (phosphohistidine intermediate) #status predicted

Query Match 37.2%; Score 613.5; DB 2; Length 432;  
Best Local Similarity 45.3%; Pred. No. 3.6e-44;  
Matches 194; Conservative 44; Mismatches 79; Indels 111; Gaps 36;

QY 2 SLLAALLAALAP-AAAAAP-LKLEKVVLSHGVRSPYKA--LM--VSPD-WP-W 52  
DB 3 AILIPFSLIPLTPQSAFAQSPPELKESVIVSRHGVRAPTKATQLMQDVTDPDAMP 62  
QY 53 PV---GLTPRGAALV-LLGY---RYFARGLL--PGCPAAG--TILADVDETR-TGQAF 99  
DB 63 PVKLGWLTFRGGELIAYLGHYQRLVADGLLTKKGCPOGQVAILADVDETRTKTGE 122  
QY 100 AAGLAPGCAI-VH-----AGDDPIFGHLDTG-C-LDQA---DAILAAGEGLT--ARH 146  
DB 123 AAGLAPDCAITVHTQADTSSPDPLFNPLKTVGCVQLDNANVTDAILSRAGGSIAFTGT 182  
QY 147 T-----LAKVLFN--ASACLE-----CARVVRGLGPI--LASTLS 179  
DB 183 TAFRELVNFPQSNLCNREKQDECSLTQALPSELKVSADNVSLTGAVSLASMLT 242  
QY 180 FLLEYAQQGMEVGVGRI--SAEW-----LLHNAQF-LMNRTPYIA--RATPI--LIVTA 228  
DB 243 FLLOQAQGMPEPGWGRITDSHQWNTLLSLHNAQFYLLQRTPEVARSRAATPLDLTKALT 302  
QY 229 -----PARVLLAGHDNTLA-LGG-LDL-WOLP-OPDETTPGGALVVELNW 271  
DB 303 PHPPKQAYGVTLPTSVLFIAGHDNTLANLGALNLWTLPGQPDNTPPGSELVFERW 362  
QY 272 -----YRVVMYQMDQLRNLPLPRILP-----IPGC---GSEAACSLSDFAR 314  
DB 363 LSDNSQWIVQSLVFTLQMDKTPSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQ 422  
QY 315 -----APAC 318  
DB 423 NEARIPAC 430

RESULT 2  
H90770  
Phosphoanhydride phosphorylase [imported] - *Escherichia coli* (strain O157:H7, substrain  
C;Species: *Escherichia coli*  
C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
C;Accession: H90770  
R;Hayashi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A;Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gene  
A;Reference number: A99629; MUID:21156231; PMID:11258796  
A;Accession: H90770  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-434 <HAY>  
A;Cross-references: GB:BA000007; PIDN:BA034559.1; PID:G13360596; GSPDB:GN00154  
A;Experimental source: strain O157:H7, substrain RMD 0509952  
C;Genetics:  
A;Gene: Ece1136

Query Match 36.8%; Score 606.5; DB 2; Length 434;  
Best Local Similarity 44.9%; Pred. No. 1.4e-43;  
Matches 193; Conservative 43; Mismatches 81; Indels 113; Gaps 36;

QY 2 SLLAALLAALAPAAAAA---EP-LKLEKVVLSHGVRSPYKA--LM--VSPD-WP 51  
DB 3 AILIPFSLIPLTPQSAFAQSPPELKESVIVSRHGVRAPTKATQLMQDVTDPDAMP 62  
QY 52 -WPV---GLTPRGAALV-LLGY---RYFARGLL--PGCPAAG--TILADVDETR-TGQ 97  
DB 63 NWPVKLGWLTFRGGELIAYLGHYQRLVADGLLTKKGCPOGQVAILADVDETRTKTGE 122  
QY 98 AFAAGLAPGCAI-VH-----AGDDPIFGHLDTG-C-LDQA---DAILAAGEGLT--ARH 144  
DB 123 AFAAGLAPDCAITVHTQADTSSPDPLFNPLKTVGCVQLDNANVTDAILSRAGGSIAFTGT 182  
QY 145 GLT-----LAKVLFN--ASACLE-----CARVVRGLGPI--LASTLS 177  
DB 183 RQAFRELVNFPQSNLCNREKQDECSLTQALPSELKVSADNVSLTGAVSLASMLT 242  
QY 178 EIFLLEYAQQGMEVGVGRI--SAEW-----LLHNAQF-LMNRTPYIA--RATPI--LIVTA 226  
DB 243 EIFLLOQAQGMPEPGWGRITDSHQWNTLLSLHNAQFYLLQRTPEVARSRAATPLDLTKALT 302  
QY 227 LS-----PARVLLAGHDNTLA-LGG-LDL-WOLP-OPDETTPGGALVVELW 269  
DB 303 LTPHPKQAYGVTLPTSVLFIAGHDNTLANLGALNLWTLPGQPDNTPPGSELVFERW 362  
QY 270 NR-----YRVVMYQMDQLRNLPLPRILP-----IPGC---GSEAACSLSDFAR 312  
DB 363 RRLSDNSQWIVQSLVFTLQMDKTPSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQ 422  
QY 313 LV-----APAC 318  
DB 423 IVNEARIPAC 432

RESULT 3

D85633

hypothetical protein appA [imported] - *Escherichia coli* (strain O157:H7, substrain EDI  
C;Species: *Escherichia coli*  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C;Accession: D85633  
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayh  
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodac  
Nature 409, 529-533, 2001  
A;Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.  
A;Reference number: A85480; MUID:21074935; PMID:11206551  
A;Accession: D85633  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-444 <STO>  
A;Cross-references: GB:AE005174; NID:G12514245; PIDN:AAG55528.1; GSPDB:GN00145; UWGP:2  
A;Experimental source: strain O157:H7, substrain EDL933  
C;Genetics:  
A;Gene: appA

Query Match 36.8%; Score 606.5; DB 2; Length 444;  
Best Local Similarity 44.9%; Pred. No. 1.4e-43;  
Matches 193; Conservative 43; Mismatches 81; Indels 113; Gaps 36;

QY 2 SLLAALLAALAPAAAAA---EP-LKLEKVVLSHGVRSPYKA--LM--VSPD-WP 51  
DB 13 AILIPFSLIPLTPQSAFAQSPPELKESVIVSRHGVRAPTKATQLMQDVTDPDAMP 72  
QY 52 -WPV---GLTPRGAALV-LLGY---RYFARGLL--PGCPAAG--TILADVDETR-TGQ 97  
DB 73 NWPVKLGWLTFRGGELIAYLGHYQRLVADGLLTKKGCPOGQVAILADVDETRTKTGE 132  
QY 98 AFAAGLAPGCAI-VH-----AGDDPIFGHLDTG-C-LDQA---DAILAAGEGLT--ARH 144  
DB 133 AFAAGLAPDCAITVHTQADTSSPDPLFNPLKTVGCVQLDNANVTDAILSRAGGSIAFTGT 192

QY 145 GLT---LAKVNF---ASACLE---CARVVGRLGPL-LASTUS 177  
Db 193 RQAFRELEEVNFPOSNLCLNREKQDSCSLTQALPSLKVSDNVSLTGAVSLASMLT 252  
QY 178 EIFLLEYAQPMVEVGR1--SAEW-----LHNAQF-LMNRTPYIA--RATPI--LIIVA 226  
Db 253 EIFLLOQAQMPGPGWGRITDSHQWNTLLSLHNAQFYLLQRTPEVARSATPLDLMITA 312  
QY 227 LS-----PARVVLLAGHDNTLA-LGG-LDL-WOLF-QPDETPPGGALVFELW 269  
Db 313 LTHPPQKQAYGVTLPFSVLFIAGHDNTLANLGALELNLWLPQPDNTPPGGELVFERW 372  
QY 270 NR-----YRVMTQMDQLNRLEPLLRIPL-----IPGC-----GSEACSLSDPAR 312  
Db 373 RRLSDNSQWITQVSLFVFTLOQMRDKTPLSLNTPPGEVYKLTACBERRNAQGMCSLAGFTQ 432  
QY 313 LV-----APAC 318  
Db 433 IVNEARIPAC 442

RESULT 4  
AC0201  
acid phosphatase (EC 3.1.3.2) [imported] - Yersinia pestis (strain CO92)  
C:Species: Yersinia pestis  
C>Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AC0201  
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Ill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, Nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AC0201  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-441 <KUR>  
A:Cross-references: GB:AL590842; PIDN:CAC90470.1; PID:g15979685; GSPDB:GN00175  
A:Gene: YPoi648  
C:Keywords: phosphoric monoester hydrolase

Query Match 35.8%; Score 590; DB 2; Length 441;  
Best Local Similarity 43.4%; Pred. No. 3.5e-42;  
Matches 187; Conservative 42; Mismatches 88; Indels 114; Gaps 34;

QY 1 VSLLLAALLAAL-APAAAAEPLKLVILSRHGVSRPTK--ALM--VSPD-WP-WP 53  
Db 10 LSGVLMLSLGLAATAPVAAEPGTYTLERVVILSRHGVSRPTKQTQLMNDVTPDRKWPQ 69  
QY 54 VG---LTPRGAALVLL-----GVRYFARGLL-PCGPAAGTIL--ADVDERTR-TGQAF 101  
Db 70 VKAGYLTPRGAELVTLMGGFVGDYFSLGILLAGCPAEGGVYACADIDQTRLTGQAF 129  
QY 102 GLAPGCAI-VH-----AGDDPIFHGLDTG-C-LD--QADAILAAGEGG-----LTARHGL 147  
Db 130 GVAPGCGLTVHNQADLKKTDPLFHFVEAGVCKLDAADKAIBEQLGGLDTVTSQYAKP 189  
QY 148 LAK---VLNFAA--CLECAR-----VVGRLGPL-LASTLSEI 179  
Db 190 FAQMGDVNLFNFAAPCYCKSLQOQGTCDFAHPAANEVNVNKEGTVKVTLSGLPALSSITLGEI 249  
QY 180 FLLEYAQPMVEVGR1--SAEWL-----LHNAQF-LMNRTPYIAR--ATPIL--IVTAL- 227  
Db 250 FLLQNAQAMPEVAMQRLKGAENWVSLSLHNAQFNLMKTPYIARHKGTPLLQOQIDTALT 309  
QY 228 -----SPARVVLLAGHDNTLA-----LGGDLWLQPL-QPDETPPGGALVFELW 269  
Db 310 LQDAQOQKLPISAQNRVLFGGHDNTIANIAGMLGAD-WOLFQPDNTPPGGELVFERW 368  
QY 270 -----NRY--VRVMTQMDQLNRLEPLLRIPL-----IPGC-----GSEACSLSDP- 310

Db 369 QNPNDHORYVAVKMFYQTMQDLNRNAEKLDLKNNPAGIISVAVAGCENNGDDKLCELDTFQ 428  
QY 311 ---ARLVAPAC 318  
Db 429 KKVAKVIEPAC 439

RESULT 5  
E87316  
periplasmic phosphoanhydride phosphohydrolase [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C>Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C:Accession: E87316  
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, C.B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolc, N.; J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.N. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: E87316  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-414 <STO>  
A:Cross-references: GB:AB005673; NID:g13421731; PIDN:AAK22529.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CC0542

Query Match 22.9%; Score 377.5; DB 2; Length 414;  
Best Local Similarity 35.9%; Pred. No. 2.9e-24;  
Matches 148; Conservative 47; Mismatches 112; Indels 105; Gaps 29;

QY 4 LLAALLAALLAALAPAAAAEPLKLVILSRHGVSRPTKALMVSPD-----WP--- 51  
Db 5 LRIAVVTLALASAGASAE--TLEKVILSRHGVSR-----AMSSPERLEEASARWPWPE 58

QY 52 WPMV-LTPRGAALV-LG--YR--YFARGLL-PCGPAAGTILADVDETR-RTGQFAAGL 103  
Db 59 VPAGHLTARGETLVARMGDYVRRHYAAQGLKPGDCASVYAWANVTQRTIATAKAYRTL 118  
QY 104 APGCAI-----VHAGD-DPIFEGHDTGCLDQADAILA-----AGE--GGTLARHGLTAKVL 152  
Db 119 APGCPVTNTVGEGNIDFEPFVKAGIV-KADHALARAANVAGRGCDLTANSASHNQAE 177  
QY 153 NFASACLECAR-----VVGRLGP-LLASTSLRIFLLEYAQG 187  
Db 178 QLDALLMQCDKPCPPAPGKRVFDKPGFVDGEBELAGLGPFAFASGVTSLLMAWADG 237  
QY 188 P--MEVHGRIAE-----WLLHNAQF-LMNRTPYIAR-----ATPILIVTALS 228  
Db 238 RDFAGLGRKSLDEEALTRSFLLHQAEFLRLRTPVARTLHAGLADRLAATLRDGAAG 297  
QY 229 P--ARVVLLAGHDNTLA-LGGL--DLWOLF-QPDETPPGGALVFELW-----RYVRM 276  
Db 298 PVDARLVIIAGHDGTLASLGGLLRMEWTLPGVQPNQIQPGALVFERWRDDGVVRVVR 357  
QY 277 Y--QTMQDLNRLEPLLRIPL-----IPGCSEAA-----CSLSDPARLVAPA 317  
Db 358 FTGQSLSLRNMTALDAKTPPLSAPFVFGCGGTATPAFDCLREDPETVVRGA 409

RESULT 6  
F90773  
periplasmic glucose-1-phosphatase [imported] - Escherichia coli (strain O157:H7, subserotype O157:H7)  
C:Species: Escherichia coli  
C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
C:Accession: F90773  
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.C.; Sasagawa, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H. DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and Serotype O157:H7  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: F90773  
A>Status: preliminary



A:Molecule type: DNA  
A:Residues: 1-413 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA34581.1; PID:g13360618; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RIND 050952  
C:Genetics:  
A:Gene: Ecs1158

Query Match 17.6%; Score 289.5; DB 2; Length 413;  
Best Local Similarity 28.6%; Pred. No. 8e-17;  
Matches 116; Conservative 71; Mismatches 116; Indels 103; Gaps 29;  
QY 1 VSLLLAALLAALAPAAAAPLEKLVILSRHVSPT-----KALMVS-----PDWP 51  
DB 6 TAAAVAGVLLASNAQAQTVEGYQLQVLMMSRNLRLAPLANNVSVLEQSTPNKWPED 65  
QY 52 WPGV-LTPRGAAL-VLLGY---RYFA-RGLLPG--CPAAGTILADVDERTR---TGOAFA 100  
DB 66 VPGQLTTKGGVLEVMGHYRWELAEQGMVKSCEPPPTVYVANSLSQRTVATAQFFI 125  
QY 101 AGLAPCAIVHAGDDPIFHGLDTGCLDQ-----ADAILAAGEGLTARHGLT 147  
DB 126 TGAFPGC-----DIPVHQRKMGTMPTFNPTVITDSDAFAEQVAAMEKEL-SKLQUT 178  
QY 148 -----LAKVLNF--ASACLE--CARVVGCR-----LGPL-LASTLSIFLLE 183  
DB 179 DSYQLLEKIVNYKDSACKKQCSIVDGKNTFSKAKYQCEPGVSGPLKVGNSLVDAFTLQ 238  
QY 184 YAGQ-PM-EVWGGRISA--EW-----LLHNAQFLMNRTPYIAR--ATPI-----LIVT 225  
DB 239 YIEGFFMDQVANGKEISDQCKVLSKLKNGYQDSLFTSPEVARNVAKPLVSYIDKALVTD 298  
QY 226 ALSPARVLLAGHDTNLA--LGGLDL--WQL-PQDPTPPGALVPFLW-----NR---Y 272  
DB 299 RTSAPKITVLVGHDSNIASLLTALDFKPYQLHDQNERITIGGKIIVQRWRDSSKANRDLMK 358  
QY 273 VRVMYTMQDLNLEPLPR-----ILPIPGC--GSEAAACSLSDP 310  
DB 359 IEYVYQSAEQLRNADALTLOAPAQRTVLELSGCPDADGFCPMDKF 404

RESULT 7  
B85636  
periplasmic glucose-1-phosphatase [imported] - Escherichia coli (strain O157:H7, substra  
C:Species: Escherichia coli  
C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C:Accession: B85636  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: B85636  
A:Molecule type: DNA  
A:Residues: 1-413 <STO>  
A:Cross-references: GB:AE005174; NID:g12514273; PIDN:AAG55550.1; GSPDB:GN00145; UNGP:214  
A:Experimental source: strain O157:H7, substrain SDL933  
C:Genetics:  
A:Gene: agp

Query Match 17.6%; Score 289.5; DB 2; Length 413;  
Best Local Similarity 28.6%; Pred. No. 8e-17;  
Matches 116; Conservative 71; Mismatches 116; Indels 103; Gaps 29;  
QY 1 VSLLLAALLAALAPAAAAPLEKLVILSRHVSPT-----KALMVS-----PDWP 51  
DB 6 TAAAVAGVLLASNAQAQTVEGYQLQVLMMSRNLRLAPLANNVSVLEQSTPNKWPED 65  
QY 52 WPGV-LTPRGAAL-VLLGY---RYFA-RGLLPG--CPAAGTILADVDERTR---TGOAFA 100  
DB 66 VPGQLTTKGGVLEVMGHYRWELAEQGMVKSCEPPPTVYVANSLSQRTVATAQFFI 125  
QY 101 AGLAPCAIVHAGDDPIFHGLDTGCLDQ-----ADAILAAGEGLTARHGLT 147

DB 126 TGAFPGC-----DIPVHQRKMGTMPTFNPTVITDSDAFAEQVAAMEKEL-SKLQUT 178  
QY 148 -----LAKVLNF--ASACLE--CARVVGCR-----LGPL-LASTLSIFLLE 183  
DB 179 DSYQLLEKIVNYKDSACKKQCSIVDGKNTFSKAKYQCEPGVSGPLKVGNSLVDAFTLQ 238  
QY 184 YAGQ-PM-EVWGGRISA--EW-----LLHNAQFLMNRTPYIAR--ATPI-----LIVT 225  
DB 239 YIEGFFMDQVANGKEISDQCKVLSKLKNGYQDSLFTSPEVARNVAKPLVSYIDKALVTD 298  
QY 226 ALSPARVLLAGHDTNLA--LGGLDL--WQL-PQDPTPPGALVPFLW-----NR---Y 272  
DB 299 RTSAPKITVLVGHDSNIASLLTALDFKPYQLHDQNERITIGGKIIVQRWRDSSKANRDLMK 358  
QY 273 VRVMYTMQDLNLEPLPR-----ILPIPGC--GSEAAACSLSDP 310  
DB 359 IEYVYQSAEQLRNADALTLOAPAQRTVLELSGCPDADGFCPMDKF 404

RESULT 8  
JV0087  
glucose-1-phosphatase (EC 3.1.3.10) precursor - Escherichia coli (strain K-12)  
C:Species: Escherichia coli  
C>Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 01-Mar-2002  
C:Accession: JV0087; H64841  
R:Pradel, E.; Marck, C.; Boquet, P.L.  
J. Bacteriol. 172, 802-807, 1990  
A:Title: Nucleotide sequence and transcriptional analysis of the Escherichia coli agp  
A:Reference number: JV0087; MUID:90130318; PMID:2153660  
A:Accession: JV0087  
A:Molecule type: DNA  
A:Residues: 1-413 <PRA>  
A:Cross-references: GB:M33807; NID:g145217; PIDN:AAA23426.1; PID:g145218  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;  
A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: H64841  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-413 <BLAT>  
A:Cross-references: GB:AE000202; GB:U00096; NID:g1787233; PIDN:AAC74087.1; PID:g178723  
A:Experimental source: strain K-12, substrain MGL655  
C:Genetics:  
A:Gene: agp  
A:Map position: 23 min  
C:Function:  
A:Description: essential for growth in a high-phosphate medium containing glucose-1-ph  
A:Note: Optimal at low pH  
C:Keywords: homodimer; periplasmic space; phosphohistidine; phosphoprotein; phosphoric  
F1-22/Domain: signal sequence #status predicted <sig>  
F123-413/Product: glucose-1-phosphatase #status predicted <MAT>  
F140/Active site: His (phosphohistidine intermediate) #status predicted  
F1311/Active site: His #status predicted

Query Match 17.5%; Score 288.5; DB 2; Length 413;  
Best Local Similarity 28.6%; Pred. No. 9.7e-17;  
Matches 116; Conservative 71; Mismatches 116; Indels 103; Gaps 29;  
QY 1 VSLLLAALLAALAPAAAAPLEKLVILSRHVSPT-----KALMVS-----PDWP 51  
DB 6 TAAAVAGVLLASNAQAQTVEGYQLQVLMMSRNLRLAPLANNVSVLEQSTPNKWPED 65  
QY 52 WPGV-LTPRGAAL-VLLGY---RYFA-RGLLPG--CPAAGTILADVDERTR---TGOAFA 100  
DB 66 VPGQLTTKGGVLEVMGHYRWELAEQGMVKSCEPPPTVYVANSLSQRTVATAQFFI 125  
QY 101 AGLAPCAIVHAGDDPIFHGLDTGCLDQ-----ADAILAAGEGLTARHGLT 147  
DB 126 TGAFPGC-----DIPVHQRKMGTMPTFNPTVITDSDAFAEQVAAMEKEL-SKLQUT 178

```

148 QY 148 -----LAKVLF--ASACLE---CARVGR-----LGPL-LASTLSEIFLLE 193
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
179 Db 179 DSYQLLEKIINYKDSPAKQKQCSLDVGKXTFSAKYQQEPGVSGPLKVGNSLVDFTLQ 238
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
184 QY 184 YAQG-PM-EVWGGRISA--EW-----LLHNAQFLMNRTPVIAR--ATPI-----LIVT 225
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
239 Db 239 YIEGPPMDQVANGEIKSDQMKVSLKNGVQDSLFTSPVARNVAKPLVSYIDKALVTD 298
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
226 QY 226 ALSGARVLVLAGHDTNIA--LGGIDL--WOL-POPDETPPGGALVPFLM-----NR---Y 272
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
299 Db 299 RTSAPKITVLVGHDSNTIASLITLDFKPYQLHDQNERTPIGGKIVFQRMHDSKANRDLMK 358
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
273 QY 273 VRVYQTMDOQLRNLPLELPR-----ILPIPGC--GSEAAACSLSDR 310
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
359 Db 359 IEVYQSAEQLRNADALTLOAPAQRVTLSEGGCPIDADGFCPMDKF 404
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
AG0632
glucose-1-phosphatase precursor (Glpase), secreted [imported] - Salmonella enterica subsp
C/Species: Salmonella enterica subsp. enterica serovar Typhi
C/Note: this species has also been called salmonella typhi
C/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C/Accession: AG0632
R.;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moutle, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.A.
A/Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A/Reference number: AB0502; MUID:21534947; PMID:11677608
A/Accession: AG0632
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-413 <PAR>
A/Cross-references: GB:AL513382; PIDN:CAN08242.1; PID:g16502289; GSPDB:GN00176
C/Genetics:
A/Gen: STY1153

```

S2527  
Glucose-1-phosphatase (EC 3.1.3.10) precursor - Providencia rettgeri  
C/Species: Providencia rettgeri  
C/Date: 25-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 08-Oct-1999  
C/Accession: S25627  
R/Riccio, M.L.; Chiesurin, A.; Lombardi, G.; Satta, G.  
submitted to the EMBL Data Library, September 1992  
A/Reference number: S25627  
A/Accession: S25627  
A/Molecule type: DNA  
A/Residues: 1-417 <R1C>  
A/Cross-references: EMBL:X69201; NID:945771; PIDN:CAA48288.1; PID:945772  
A/Experimental source: strain PV7  
C/Genetics:  
A/Gene: agp  
C/Keywords: periplasmic space; phosphohistidine; phosphoprotein; phosphoric  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:30-417/Product: glucose-1-phosphatase #status predicted <WAT>  
F:42/Active site: His (phosphohistidine intermediate) #status predicted  
F:314/Active site: His #status predicted

Query Match 15.5%; Score 256; DB 2; Length 417;  
Best Local Similarity 29.0%; Pred. No. 5.5e-14;  
Matches 119; Conservative 57; Mismatches 140; Indels 94; Gaps

Qy 1 VSLIAALALAAAPAAAAA--PLKLEKVVILSRHGVRSPP--TKALMVS-----PD  
Db 6 LTLCUSAAL--FAPIAPTWASDINQADWLDQVLVLSRHNRITFVNTGILTEVTDKKWPD  
Qy 50 WMPVVG-LTPRGAALVLLGYRFARGLLPG-----CPAAGT--ILADVDETR-RTQG  
Db 65 WDAKSGYLTTQGALEVMVGHYFREWIDQNKLLADELCPTSNEDIVLYTNSLQRTIATAQ  
Qy 98 AFAAGLAPCAI-VH-----AGDPIPHGLDT-CGLDQADAILAAGSG--GLTASHGL-  
Db 125 FPAAGAPPOCKVNIHQPEIGWMDVFNPITNGSPFEKQKALAAAMDDYLLKGLSLKAGYE  
Qy 147 TLAKYLNFA-----SACLECARVVVRIGPL-LASTLSEIFLLEVAQG  
Db 185 ELDTVLINIKDSQCKTDXLCLNLDQKNSFIIEADKEFGVSGPLKIANSAVDAILDQVYEG  
Qy 188 --PMEVWGCRISA--EW---LLHNAQPLMNRTPVIAR--ATPIL-----IVTALS  
Db 245 FPADQVAVGLVDTPEKWKKLNTLKNVQETLFTPKIATKNVAHPILNYIDKGFVSVDKGE  
Qy 229 PARVILLAGHTNLA-LGGDL-LWLPQP-DETPPGGALVFELMN-----RYVRV  
Db 305 TAKFIFLVGHDSNIASLMSAMDFKPYQAAQYEHPTIGGKLIVFORWTDKTKDFMKVEY  
Qy 276 MYQTMQLR-----NLE-PLLPRIILIPGC--GSEAAACSLSDPARLVAPA 317  
Db 365 VYTADQLRDNAYLSETPPKHVILELXDCPVDKNGVCSWEDFKVWAKA 414

RESULT 11  
A33395  
acid phosphatase (EC 3.1.3.2) precursor - rat  
C/Species: Rattus norvegicus (Norway rat)  
C/Date: 29-Jan-1990 #sequence\_revision 29-Jan-1990 #text\_change 22-Jun-1999  
C/Accession: A33395  
R/Himeno, M.; Fujita, H.; Noguchi, Y.; Kono, A.; Kato, K.  
Biochem. Biophys. Res. Commun. 162, 1044-1053, 1989  
A/Title: Isolation and sequencing of a cDNA clone encoding acid phosphatase  
A/Reference number: A33395; MUID:89350910; PMID:2764916  
A/Accession: A33395  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-423 <HIM>  
A/Cross-references: GB:M27893; NID:9202933; PIDN:AAA40744.1; PID:G9202934  
C/Superfamily: mammalian acid phosphatase  
C/Keywords: phosphohistidine; phosphoprotein; phosphoric monoester hydrolase  
F:41/Active site: Arg #status predicted  
F:42/Active site: His (phosphohistidine intermediate) #status predicted

Query Match 8.2%; Score 134.5; DB 2; Length 423;  
Best Local Similarity 25.3%; Pred. No. 0.001;  
Matches 95; Conservative 33; Mismatches 157; Indels 91; Gaps 20;

QY 4 LLAALLAALAPAAAAEPLKLVILSRHGVSPTKALMVSP--DWPWVG---LTP 58  
DB 13 LLQFLGLMCLVMPPIQARSLF--VTLRYHGRDSPVKAYPKDPYQBEKWPQGGQUTK 70  
QY 59 RGAAL-----VLGVRYPARGLLPGCPAAGTTL--ADVDERTRTGOAFAAGLAPGCAIV 110  
DB 71 EGMLOHWELOALRORY--HGFLNASHYRQEVYVSTDFDRTLSAEANLAGLPPPTVEQ 128  
QY 111 H-----AGDDPIFHGLDTG-----CLDQADAI-L 133  
DB 129 HFNPNISWQIPVHTVPTEDRLKLPPLGCPCEPRYEQLONETROTPEYQNMSTQNAQFLDM 188  
QY 134 RAGEGLTARHGLTAKLVNFASACLECARVVRGLGPLLAS-----TLSEI---FLLEY 184  
DB 189 VANETGLM--NLTLETINWVDT-LFCEQTHGLLLPWASPTQVQALSQKDSFPLF 244  
QY 185 AQGPMEVWGGRISAEWLLHNAQFLMNRTPYIARATPILIVTALS PARVLLAGHDTNLAL 244  
DB 245 GIHD-QVKARLQGGVLL--AQILKNLT-----LMATTSQPPKLLVYSAHDTTLVA 292  
QY 245 GGLDLWLPQDETTPGGALVFELW---NRVYRVYQTMQDLNLEPLPRILPIPGCG 300  
DB 293 LQMAL-NVYNGKQAPYASCHIFELYQEDNGNFSVEMY-----FRNDSKAPWFLRLTEP 346  
QY 301 SEACSLSDPARLVAP 316  
DB 347 HR--CPLQDFLRLETP 360

## RESULT 12

S14742  
acid phosphatase (EC 3.1.3.2) precursor - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C>Date: 21-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 31-Mar-2000  
C:Accession: S14742  
R:Geier, C.; von Figura, K.; Pohlmann, R.  
Biol. Chem. Hoppe-Seyler 372, 301-304, 1991  
A:Title: Molecular cloning of the mouse lysosomal acid phosphatase.  
A:Reference number: S14742; MUID:91282986; PMID:2059337  
A:Accession: S14742  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-421 <GEI>  
A:Cross-references: EMBL:X57199; NID:G52870; PIDN:CAA40485.1; PID:G52871  
C:Superfamily: mammalian acid phosphatase  
C:Keywords: lysosome; phosphoric monoester hydrolase

Query Match 7.7%; Score 127.5; DB 2; Length 421;  
Best Local Similarity 23.5%; Pred. No. 0.004;  
Matches 92; Conservative 32; Mismatches 145; Indels 123; Gaps 19;

QY 6 AALL-----ALAAPAAAAEPLKLVILSRHGVSPTKALMVSP-----DWP----- 51  
DB 9 AALLQFLGLMCLVMPPIQARS-----LRFVTLRYHGRDSPVKTYPKDPYQBEKWPQGGF 64  
QY 52 -----WPVGLTRPGAALVLLGYRY-----FARGLLPGCPAAGTILADVDERTRTG 96  
DB 65 QLTKEGMLQHWELQALRORYHGFNTSYRQEVYRS-----TDFDRLTMSA 112  
QY 97 QAFAGLAPGCAIVH-----AGDDPIFHGL----- 121  
DB 113 EAMLAGLFPPEVNEVHFSFNISWQIPVHTVPTEDRLKLPPLGCPCEPRYEQLONETROTPE 172  
QY 122 -DTGCLQADAI-LAAGEGGLTARHGLTAKLVNFASACLECARVVRGLGPLLAS----- 174  
DB 173 YQNSRISQAGFLNVMANETGLT---NLTLETINWVDT-LFCEQTHGLLLPWASPTQVQ 228  
QY 175 -----TLSEIFLLEYAQGPMEVWGGRISAEWLLHNAQFLMNRTPYIARATPILIVTALS 228

DB 229 RLSQKDFSLFLFGIHE---QVKARLQGGVLL--AQILKNLT-----LMATTSQ 274  
QY 229 PARVLLAGHDTNLALGGLDLWLPQDETTPGGALVFELW---NRVYRVYQTMQDLR 284  
DB 275 FKLLVYSAHDTTLVALQMAL-NVYNGKQAPYASCHIFELYQEDNGNFSVEMY-----FR 328  
QY 285 NLSPFLPRILPIPGCGSEACSLSDPARLVAP 316  
DB 329 NDSKAPWFLRLTEP 358

## RESULT 13

S06167  
acid phosphatase (EC 3.1.3.2) ACP2 precursor [validated] - human  
N:Alternate names: acid phosphatase, lysosomal  
C:Species: Homo sapiens (man)  
C>Date: 04-Dec-1992 #sequence\_revision 01-Dec-1995 #text\_change 08-Dec-2000  
C:Accession: S06167; S05525; S01155  
R:Geier, C.; von Figura, K.  
submitted to the EMBL Data Library, June 1989  
A:Reference number: S06167  
A:Accession: S06167  
A:Molecule type: DNA  
A:Residues: 1-423 <VON>  
A:Cross-references: EMBL:X15525; NID:G34239; PIDN:CAA33542.1; PID:G1199524  
R:Geier, C.; von Figura, K.; Pohlmann, R.  
Eur. J. Biochem. 183, 611-616, 1989  
A:Title: Structure of the human lysosomal acid phosphatase gene.  
A:Reference number: S05525; MUID:89377828; PMID:2776754  
A:Accession: S05525  
A:Molecule type: DNA  
A:Residues: 1-29 <GEI>  
R:Pohlmann, R.; Krentler, C.; Schmidt, B.; Schroeder, W.; Lorkowski, G.; Culley, J.; M  
EMBO J. 7, 2343-2350, 1988  
A:Title: Human lysosomal acid phosphatase: cloning, expression and chromosomal assign  
A:Reference number: S01155; MUID:89052645; PMID:3191910  
A:Accession: S01155  
A:Molecule type: mRNA  
A:Residues: 1-423 <POH>  
A:Cross-references: EMBL:X12548; NID:G34262; PIDN:CAA31064.1; PID:G34263  
A:Note: part of this sequence, including the amino end of the mature protein, was conf  
C:GeneticS:  
A:Gene: GDB:ACP2  
A:Cross-references: GDB:118963; OMIM:171650  
A:Map position: 11p11.2-11p11.11  
A:Introns: 38/3; 70/3; 99/3; 150/3; 183/3; 213/3; 259/1; 285/3; 321/2; 380/1  
C:Function:  
A:Description: catalyzes the hydrolysis of a wide range of phosphate esters  
C:Superfamily: mammalian acid phosphatase  
C:Keywords: glycoprotein; phosphonitidine; phosphoprotein; phosphoric monoester hydrol  
F:1-30/Domain: signal sequence #status predicted <SIG>  
F:31-423/Product: acid phosphatase ACP2 #status experimental <MAT>  
F:41/Active site: Arg #status predicted  
F:42/Active site: His (phosphohistidine intermediate) #status predicted  
F:92,133,167,177,191,267,322,331/Binding site: carboxylate (Asn) (covalent) #status f  
F:159-370,212-310,345-349/Disulfide bonds: #status predicted

Query Match 7.6%; Score 124.5; DB 1; Length 423;  
Best Local Similarity 23.9%; Pred. No. 0.0071;  
Matches 95; Conservative 32; Mismatches 136; Indels 135; Gaps 23;

QY 4 LLAALLA-LAALAPAAAAEPLKLVILSRHGVSPTKALMVSP-----DWP----- 51  
DB 13 LLQLLGLVNLVMPPTARS-----LRFVTLRYHGRDSPVKTYPKDPYQBEKWPQGGQ 68  
QY 52 -----WPVGLTRPGAALVLLGYRY-----FARGLLPGCPAAGTILADVDERTRTGOA 98  
DB 69 TKEGMLQHWELQALRORYHGFNTSYRQEVYRS-----TDFDRLTMSAEA 116  
QY 99 FAAGLAPGCAI-----VHA-----GDDPIFHGL----- 121  
DB 117 NLAGLFPPEVNGMORFNPNISWQIPVHTVPTEDRLKLPPLGCPCEPRYEQLONETROTPEYQ 176

QY 122 -----DTGCLDQADAILAAGEGGLTARHGLTILAKVNFASACLECARVVG-RIGPLLA-S 174  
Db 177 NERRNAQFLD-----MWANETGLT--DLLETWVNVYDT-LFCEQTHGLRLFPWASPO 227  
QY 175 TLSEIFLLE-----YAQGPMEVGMGRISAEMLLHNAQFLMRTPIYIARATPILIV 224  
Db 228 TMQRLSKLDFSRFLFGIYQAQEK-----ARLGGVLL--AQIRKILT-----LMA 272  
QY 225 TALSPARVLLAGHDNTLAL--CGLDLWQLPQDPTPGGALVFEELW-----NRYVRVMYQ 278  
Db 273 TTSQPLKLLVYSHDITLVALQWALDVIY---NCEQAPYASCHIFELYQEDSGNFSVEMY- 328  
QY 279 TMDQLNLEPLLRILPIPGCGSEAAACSLSDFAFLVAP 316  
Db 329 ----FRNESDKAPWLSLPGCPHR--CPLQDFLRLTEP 360

RESULT 14  
E83353  
xylulose kinase PA2343 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 09-Nov-2001  
R:Accession: E83353  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Boman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lam, J.; Lory, S.; Olson, M.V.  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen  
A:Reference number: A82250; MUID:2043737; PMID:10984043  
A:Accession: E83353  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-502 <STO>  
A:Cross-references: GB:AF004660; GB:AE004091; NID:99948372; PIDN:AAG05731.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: mtly; PA2343  
C:Superfamily: xylulokinase

Query Match 7.1%; Score 117; DB 2; Length 502;  
Best Local Similarity 27.0%; Pred. No. 0.037;  
Matches 77; Conservative 24; Mismatches 118; Indels 66; Gaps 12;  
QY 67 QY-----RYFARGLLPGCPAAGTILADVDERTRTG-----QAFAGLAPGCAIVHA 112  
Db 190 GYFVRRRTWASDVLELVPEGGRILAAALPELIEFCAGICGNLRPEAAALGLAPHTRVAGG 249  
QY 113 GDDPIFHGLDTG-----CLDQADAILAAGEGGLTARHG-----LTLAKV 151  
Db 250 GGDNMALAAIGTGNIRPGLLTASLTSGTSLAYASRPLVSPHGEATFCASSGGWLPACT 309  
QY 152 LNFASACLECARVVGRLGFLPILASTLSIFLLEYAOGMEVGMGRISAEMLLHNAQFLMNR 211  
Db 310 NNLTGAC-----GLVQDLHLHLDDEFSRLA-AQAP-----VGAEGLLMLPFFDGER 354  
QY 212 TPYIARATPILIT-VTALSFAVVLLAGHDNTNLALG---GLDLWQ---LPQDETPPGGAL 264  
Db 355 VPALPHASASLHGWTAAANLSRANLCRAVLEGTAFGLRYGLDRLASGLPGEIRLNGAA 414  
QY 265 VFELNRYVRVMYQMDQLNLEPLLRILPIPGCGS--EAACSL 307  
Db 415 KNPLMRR-----TLADLLGLPLVCPROTEAAALGAALQAANSL 452

RESULT 15  
H70872  
probable ctpD protein - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
R:Accession: H70872  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: H70872  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-657 <COL>  
A:Cross-references: GB:AL021184; GB:AL123456; NID:G3261498; PIDN:CAA15997.1; PID:G27914  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: ctpD  
C:Superfamily: Enterococcus copper-transporting ATPase copB; ATPase nucleotide-binding  
F:25-394/Domain: ATPase transduction domain homology <ATP>  
F:465-608/Domain: ATPase nucleotide-binding domain homology <ATN>

Query Match 6.9%; Score 113.5; DB 2; Length 657;  
Best Local Similarity 20.4%; Pred. No. 0.099;  
Matches 80; Conservative 50; Mismatches 125; Indels 137; Gaps 17;  
QY 8 LLAALAAPAAAAEPLKLEKVVIL-----SRHGVS-----PTKALMVSP 48  
Db 96 LMTAAAVGAVAIQGFIDGALLIVFATSGALDDIATHTAESVKGLLDLAPDQAVVVG 155  
QY 49 DWPMPVGLTPRGAALVLLGYRYFARGLLPG--CPAAGTIL---ADVDERTRTGOFAAGL 103  
Db 156 DGSERV-----VAASELVVGDVVVR---PGDRIADGAVLSGASDVQDSITGESMPVAK 208  
QY 104 APGCAIVHAGDDPIFHGLDTG-----CLDQADAILA-----AGEGGLTARHGLTIL 149  
Db 209 ARG-----DEVFAGTVNGSGVLHLYVTRDPSTQTVVARIVELVADASATKATQLEFIE 260  
QY 150 KV-----LNFASACLECARVVGRLGFLPILASTL 176  
Db 261 KIEQRYSLGMVAATLALIVILMFCADLRPVLLRAMTENVIVASPCAVVLTMPPLLSAIA 320  
QY 177 SEIFLLEYAOGMEVGMGR-----ISAEMLLHNAQFLMNRTPYIARATPILIVTALS 228  
Db 321 NA-----GRHGVLKSAVVVERLADTSIVALDKTGLTRGIP----- 357  
QY 229 PARVLLAGHDNTNLALGGLDLWQL-----LPQDETPPGGALVFEELNRYVRVMYQMDQLR 284  
Db 358 --RLASVAPLDPNV-VDARRLLQALAAAEQSEHPLGRAIVAEARRRGIAI----- 405  
QY 285 NLEPLLRILPIPGCGSEAAACSLSDFAFLVAP 316  
Db 406 ---PPAKDFRAVPGCGVHALVG-NDFVEIASP 433

Search completed: April 30, 2004, 12:42:22  
Job time : 13.6341 secs

OM protein - protein search, using sw model

Run on: April 30, 2004, 12:31:10 ; Search time 7.7561 Seconds  
(without alignments)  
2134.876 Million cell updates/sec

Title: US-10-021-723B-14  
Perfect score: 1647  
Sequence: 1 VSLLLAALLALALAPAAA.....CGSEAAACSLDFARLVAPAC 318

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues  
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	613.5	37.2	432	1	PPA_ECOLI
2	288.5	17.5	413	1	AGP_ECOLI
3	284.5	17.3	413	1	AGP_SALTY
4	256	15.5	417	1	AGP_PRORE
5	134.5	8.2	423	1	PPAL_RAT
6	127.5	7.7	423	1	PPAL_MOUSE
7	124.5	7.6	423	1	PPAL_HUMAN
8	113.5	6.9	657	1	CTPD_MYCTU
9	103.5	6.3	386	1	PPAP_HUMAN
10	102.5	6.2	312	1	FMT_BORER
11	102.5	6.2	312	1	FMT_BORFA
12	100.5	6.1	411	1	YOH2_STRCO
13	98.5	6.0	492	1	MOT3_RAT
14	98	6.0	402	1	KAS2_STRCM
15	94	5.7	1307	1	BCC4_ACEXY
16	92	5.6	1098	1	EMBB_MYCTU
17	91.5	5.6	245	1	TRMD_CAUCR
18	91	5.5	772	1	METE_STRCO
19	90.5	5.5	352	1	TPO_CANFA
20	90	5.5	669	1	PP1B_STRCO
21	89.5	5.4	312	1	FMT_BORPE
22	89	5.4	402	1	DMX_SYNLE
23	88.5	5.4	432	1	MESJ_ECOLI
24	88.5	5.4	449	1	AROA_PSES2
25	88.5	5.4	450	1	AROA_BRUSE
26	88.5	5.4	450	1	AROA_BRUSU
27	88	5.3	776	1	HYPF_AZOVI
28	88	5.3	3011	1	POLG_HCVH
29	87.5	5.3	640	1	Y051_MYCTU
30	87.5	5.3	1164	1	YAV1_XANCV
31	87.5	5.3	2554	1	7LES_DRONE
32	87	5.3	375	1	COBT_STRAW
33	87	5.3	405	1	Y501_PYRAB

34 87 5.3 635 1 VP40\_HSV11 P10210 herpes simp  
35 87 5.3 1008 1 MMLA\_MYCLE Q49619 mycobacteri  
36 86.5 5.3 382 1 NORW\_VIBVU Q844f7 vibrio vuln  
37 86.5 5.3 503 1 PUR8\_STRLP P42670 streptomyce  
38 86.5 5.3 623 1 GLMS\_MYCTU O06253 m glucosami  
39 86 5.2 341 1 FAH1\_DEIRA Q9t223 deinococcus  
40 85.5 5.2 450 1 AROA\_BRUAB Q9gsv2 brucella ab  
41 85 5.2 228 1 BIOD\_PSEAB Q91614 pseudomonas  
42 85 5.2 405 1 CHSY\_EQUAR Q9mbb1 equisetum a  
43 85 5.2 654 1 SPH2\_HUMAN Q9nra0 homo sapien  
44 85 5.2 836 1 VG26\_BPM2 O64220 mycobacteri  
45 84.5 5.1 418 1 CLCB\_ECOLI P76175 escherichia

#### ALIGNMENTS

RESULT 1  
PPA\_ECOLI  
ID\_PPA\_ECOLI STANDARD; PRT; 432 AA.  
AC P07102;  
DT 01-APR-1988 (Rel. 07, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Periplasmic appa protein precursor [Includes: Phosphoanhydride  
DE phosphohydrolase (EC 3.1.3.2) (pH 2.5 acid phosphatase) (AP); 4-  
DE phytase (EC 3.1.3.26)].  
GN APPA OR B0980.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-29.  
RC STRAIN=K12;  
RX MEDLINE=90368616; PubMed=2168385;  
RA Dassa J., Marck C., Boquet P.L.;  
RT "The complete nucleotide sequence of the Escherichia coli gene appA  
RT reveals significant homology between pH 2.5 acid phosphatase and  
RT glucose-1-phosphatase."  
RL J. Bacteriol. 172:5497-5500(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12."  
RL Science 277:1453-1474(1997).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97061202; PubMed=8905232;  
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
RA Yano M., Horiuchi T.;  
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome  
RT corresponding to the 12,7-28.0 min region on the linkage map."  
RL DNA Res. 3:1137-1155(1996).  
RN [4]  
RP SEQUENCE OF 1-112 FROM N.A.  
RX MEDLINE=87271766; PubMed=3038201;  
RA Touati B., Danchin A.;  
RT "The structure of the promoter and amino terminal region of the pH  
RT 2.5 acid phosphatase structural gene (appA) of E. coli: a negative  
RT control of transcription mediated by cyclic AMP."  
RL Biochimie 69:215-221(1987).  
RN [5]

RC SEQUENCE OF 1-17 FROM N.A.  
 RC STRAIN=K12;  
 RA MEDLINE=92049231; PubMed=1658595;  
 RA Dassa J., Feih H., Marck C., Dion M., Kieffer-Bontemps M.,  
 RA Boquet P.L.;  
 RA "A new oxygen-regulated operon in *Escherichia coli* comprises the  
 RT genes for a putative third cytochrome oxidase and for pH 2.5 acid  
 RT phosphatase (appa).";  
 RL Mol. Gen. Genet. 229:341-352(1991).  
 RN [6]  
 RP CHARACTERIZATION, AND SEQUENCE OF 23-34.  
 RA Greiner R., Jany K.D.;  
 RL "Characterization of a phytase from *Escherichia coli*.";  
 RL Biol. Chem. Hoppe-Seyler 372:664-665(1991).  
 RN [7]  
 RP CHARACTERIZATION, AND SEQUENCE OF 23-35.  
 RA Greiner R., Konietzny U., Jany K.D.;  
 RA "Purification and characterization of two phytases from *Escherichia coli*.";  
 RT Arch. Biochem. Biophys. 303:107-113(1993).  
 RL [8]  
 RP CHARACTERIZATION.  
 RA MEDLINE=20161462; PubMed=10696472;  
 RA Golovan S., Wang G., Zhang J., Forsberg C.W.;  
 RA "Characterization and overproduction of the *Escherichia coli* appa  
 RT encoded bifunctional enzyme that exhibits both phytase and acid  
 RT phosphatase activities.";  
 RL Can. J. Microbiol. 46:59-71(2000).  
 RN [9]  
 RP MUTAGENESIS.  
 RA MEDLINE=93054596; PubMed=1429631;  
 RA Ostain K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.-M.,  
 RA van Etten R.L.;  
 RA "Overexpression, site-directed mutagenesis, and mechanism of  
 RT *Escherichia coli* acid phosphatase.";  
 RL J. Biol. Chem. 267:22830-22836(1992).  
 RN [10]  
 RP X-RAY CRYSTALLOGRAPHY (2.05 ANGSTROMS).  
 RA MEDLINE=20122624; PubMed=10655611;  
 RA Lim D., Golovan S., Forsberg C.W., Jia Z.;  
 RA "Crystal structures of *Escherichia coli* phytase and its complex with  
 RT pyruvate.";  
 RL Nat. Struct. Biol. 7:108-113(2000).  
 CC -1- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an  
 CC alcohol + phosphate.  
 CC -1- CATALYTIC ACTIVITY: Myo-inositol hexakisphosphate + H(2)O = 1D-  
 CC myo-inositol 1,2,3,4,5-pentakisphosphate + phosphate.  
 CC -1- SUBUNIT: Monomer.  
 CC -1- SUBCELLULAR LOCATION: Periplasmic.  
 CC -1- INDUCTION: IN ADDITION TO CAMP-MEDIATED CONTROL, THIS ENZYME IS  
 CC INDUCED WHEN BACTERIAL CULTURES REACH STATIONARY PHASE; ITS  
 CC SYNTHESIS IS TRIGGERED BY PHOSPHATE STARVATION OR A SHIFT FROM  
 CC AEROBIC TO ANAEROBIC CONDITIONS.  
 CC -1- SIMILARITY: Belongs to the histidine acid phosphatase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M58708; AAA72086.1; -;  
 DR EMBL; AE000200; AAC74065.1; -;  
 DR EMBL; D90735; BAA35745.1; -;  
 DR EMBL; X05471; CAA29031.1; -;  
 DR EMBL; M58708; -; NOT ANNOTATED\_CDS.  
 DR EMBL; S63811; AAB20286.1; -;  
 DR PIR; B36733; B36733.  
 DR PDB; 1DKL; 03-AUG-00.  
 DR PDB; 1DKM; 02-AUG-00.

DR PDB; 1DKM; 03-AUG-00.  
 DR PDB; 1DKO; 03-AUG-00.  
 DR PDB; 1DKP; 03-AUG-00.  
 DR PDB; 1DKQ; 03-AUG-00.  
 DR SWISS-2DPAGE; P07102; COLI.  
 DR EcoGene; EGI0049; appa.  
 DR InterPro; IPR000560; Hisac\_phsphtse.  
 DR Pfam; PF00328; acid\_phosphat; 1.  
 DR PROSITE; PS00616; HIS\_ACID\_PHOSPHAT\_1; 1.  
 DR PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.  
 DR Hydrolase; Periplasmic; Signal; Multifunctional enzyme; 3D-structure;  
 KW Complete proteome.  
 FT SIGNAL 1 22  
 FT CHAIN 23 432 PERIPLASMIC APPA PROTEIN.  
 FT ACT\_SITE 39 39 NUCLEOPHILIC ACCEPTOR.  
 FT ACT\_SITE 326 326 PROTON DONOR.  
 FT DISULFID 99 130  
 FT DISULFID 155 430  
 FT DISULFID 200 210  
 FT DISULFID 404 413  
 FT CONFLICT 51 66 MDVTPDAWPTWPKL -> NAGCHPRRANLAGKT (IN  
 FT REF. 3).  
 FT CONFLICT 75 76 EL -> DV (IN REF. 4).  
 FT CONFLICT 112 112 D -> S (IN REF. 4).  
 FT STRAND 28 38  
 FT STRAND 42 42  
 FT HELIX 49 53  
 FT HELIX 54 54  
 FT TURN 66 67  
 FT STRAND 69 69  
 FT HELIX 71 90  
 FT TURN 91 92  
 FT TURN 102 103  
 FT STRAND 105 109  
 FT HELIX 113 126  
 FT TURN 128 129  
 FT STRAND 134 135  
 FT TURN 140 141  
 FT HELIX 145 147  
 FT TURN 149 153  
 FT HELIX 159 169  
 FT TURN 170 171  
 FT HELIX 174 179  
 FT TURN 180 181  
 FT HELIX 182 192  
 FT TURN 193 193  
 FT HELIX 194 196  
 FT HELIX 198 201  
 FT TURN 202 202  
 FT HELIX 212 215  
 FT STRAND 220 222  
 FT STRAND 227 229  
 FT HELIX 231 248  
 FT TURN 249 250  
 FT HELIX 254 257  
 FT TURN 258 258  
 FT HELIX 262 279  
 FT TURN 280 281  
 FT HELIX 283 289  
 FT TURN 290 290  
 FT HELIX 291 301  
 FT STRAND 307 308  
 FT HELIX 310 312  
 FT STRAND 314 315  
 FT STRAND 319 324  
 FT HELIX 326 336  
 FT TURN 337 337  
 FT STRAND 342 342  
 FT TURN 343 344  
 FT STRAND 345 345  
 FT TURN 351 352  
 FT STRAND 354 362  
 FT TURN 363 366

Query Match	17.5%;	Score 288.5;	DB 1;	Length 413;
Best Local Similarity	28.6%;	Pred. No. 3.2e-15;		
Matches 116;	Conservative	71;	Mismatches 116;	Indels 103; Gaps 29;
Qy	1	VSLLAALLAALAPAA	AAAEFLKLEKVVILSRHGVR	SPT----KALMVS-----PDWP 51
Db				
Qy	6	IAAAVAGIVLLASNAQ	AVTPGEGYQLQVLMMSRNLRA	PLANNGSVLQESTPKNKWPED 65
Db				
Qy	52	WPVG-LTPRGAAL-VLLG	V---RYFA-RGLLPG---CPAAGT	ILADVDETR--TCQAF 100
Db				
Qy	66	VFGQLTTKGVVLVGYM	GHYEMLEWLAECQMWKSGEC	PPYVYVYANSLQRTVATAQFFI 125
Db				
Qy	101	AGLAPCAIVHAGDDPI	FHGLDTGCLDQ-----ADAILA	AGEGGLTASHGLT 147
Db				
Qy	126	TGAFFGC-----DI	EVHHEQKMGTMPTFN	VITDSDAAPSEQAVAMEKEL-SKLQLT 178
Db				
Qy	148	-----LAKVLNF--	ASACLE---CARVGR-----	LGPI-LASTLSETFLLE 183
Db				



```
Db 179 DSYQLEKIVNYKDSACKKQOCSLVGKNTFSKAYQOEPGVSGPLKVGNSLVDAFTLQ 238
QY 184 YAQG-PM-EVWGRI--EW-----LLHNAQFLMRTPIYAR--ATPI-----LIVT 225
Db 239 YVEGFPDQVANGKESDOQWVLSKXNGYQDSIFTSPTSVARNVAKPLVSYDKALVTD 298
QY 226 ALSPARVLLAGHDNLA--LGLDL--WQL-PQDETTPGALVFELW-----NR---Y 272
Db 299 RTSAPKITVLVGHDSNIALSLTALDFKPYQLHQDQERTPIGKIVFQRMHDSKANRDLMK 358
QY 273 VRVMYQTMQDLNLEPLLR-----ILPIQC--GSEAACSLSDF 310
Db 359 IYVYQSAEQLNADALTLQAPQAVTLELSGCPIDADGFCPMDKF 404

RESULT 3
AGP_SALTY
ID AGP_SALTY STANDARD; PRT; 413 AA.
AC O33921;
DT 15-JUL-1998 (Rel. 36, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glucose-1-phosphatase precursor (EC 3.1.3.10) (G1Pase).
GN AGP OR STM117.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856(2001).
RN [2]
RP SEQUENCE OF 1-249 FROM N.A.
RC STRAIN=TN1379;
RX MEDLINE=97405892; PubMed=9260936;
RA Gupta S.D., Wu H.C., Rick P.D.;
RT "A Salmonella typhimurium genetic locus which confers copper
RT tolerance on copper-sensitive mutants of Escherichia coli.";
RL J. Bacteriol. 179:4977-4984(1997).
CC -|- CATALYTIC ACTIVITY: D-glucose 1-phosphate + H(2)O = D-glucose +
CC phosphate.
CC -|- SUBUNIT: Homodimer (By similarity).
CC -|- SUBCELLULAR LOCATION: Periplasmic.
CC -|- SIMILARITY: Belongs to the histidine acid phosphatase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; A3008748; AAL20049.1; -.
CC DR EMBL; U75949; AAC45604.1; -.
CC DR HSPF; P07102; IDKL.
CC DR KSSP; SG10595; agp.
CC DR InterPro; IPR000560; HisAc_phsphtase.
CC DR Pfam; PF00328; acid phosphat; 1.
CC DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
CC DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
CC DR Hydrolase; Periplasmic; Signal; Complete proteome.
CC SIGNAL 1 22 BY SIMILARITY.
CC CHAIN 23 413 GLUCOSE-1-PHOSPHATASE.
```

```
FT ACT_SITE 40 40 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).
FT ACT_SITE 311 311 PROTON DONOR (BY SIMILARITY).
FT CONFLICT 57 57 T -> P (IN REF. 2).
FT CONFLICT 60 63 AWPA -> TCPP (IN REF. 2).
FT CONFLICT 74 74 K -> T (IN REF. 2).
SQ SEQUENCE 413 AA; 45559 MW; E057667CF8A7244C CRC64;

Query Match 17.3%; Score 284.5; DB 1; Length 413;
Best Local Similarity 30.8%; Pred. No. 6.5e-15;
Matches 126; Conservative 59; Mismatches 127; Indels 97; Gaps 29;

QY 4 LLAALLAALAPAAAAA---EPLKLEKWLILSRHGVRSPT---KALMVS-----PDW 50
Db 5 LLAVAVAGVLLSSAVQRTTPEGYLOQLVMSRHLRPLANNGLSVLAQSTPNAPAW 64
QY 51 FWPVG--LTPRGAAL-VLLGYR---YFARGLLPG--CPAAGTILADVDERTR---TGOAF 99
Db 65 DVEGGQLTTKGGVLEVINGHYTREMLVAQGLIPSCSPADPTVYAYANSLQRTVTAQFF 124
QY 100 AAGLAPGCAI-VHAGD-----DPIFHGLDTGCLDOADAILAAGEGGLTARHGI----- 146
Db 125 ITGAFFGCDIPVHHQKMGMTDPTFNPVTD--DSAPFQQAVQAMEKARSQHLHDES YK 182
QY 147 TLAKVLNF--ASACLE--CARV-----VGRIGPL--LASTLSEIPLLEYAOG 187
Db 183 LLEQITHYQDSPCKEKHOCSLIDAKDTFSANYQOEPGVQGLKVGNSLVDAFTLQY YEG 242
QY 188 -PM-EVWGRI--SAEW-----LLHNAQFLMRTPIYAR--ATP-----LIVTALSP 229
Db 243 FPDQVANGGIHTRQWVLSKXNGYQDSIFTSPTSVARNVAKPLVSYDKALVDRVSA 302
QY 230 ARVVLVLAGHDNLA--LGLDL--WQL-PQDETTPGALVFELW-----NR---YVRVM 276
Db 303 PKVTVLVGHDSNIALSLTALDFKPYQLHQDQERTPIGKIVFQRMHDSKANRDLMKI YV 362
QY 277 YQTMQDLNLEPLLR-----ILPIPGGSEA--ACSLSDARLAVAPA 317
Db 363 YQSARQLRNBALTLKSPAQRVTLELKGCPVDANGFCPLDKFDNVNNTA 411

RESULT 4
AGP_PROBE
ID AGP_PROBE STANDARD; PRT; 417 AA.
AC Q52309;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Glucose-1-phosphatase precursor (EC 3.1.3.10) (G1Pase).
GN AGP.
OS Providencia rettgeri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Providencia.
OX NCBI_TaxID=587;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PV7;
RA Riccio M.L., Chiesurin A., Lombardi G., Satta G.;
RL Submitted (SEP-1992) to the EMBL/GenBank/DBJ databases.
CC -|- CATALYTIC ACTIVITY: D-glucose 1-phosphate + H(2)O = D-glucose +
CC phosphate.
CC -|- SUBUNIT: Homodimer (By similarity).
CC -|- SUBCELLULAR LOCATION: Periplasmic (By similarity).
CC -|- SIMILARITY: Belongs to the histidine acid phosphatase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X68201; CAA48288.1; -.
CC DR
```

DR PIR; S25627; S25627.  
DR HSP; P07102; 1DKL.  
DR InterPro; IPR000560; HisAc\_phosphatase.  
DR Pfam; PF00328; acid\_phosphat; 1.  
DR PROSITE; PS00616; HIS\_ACID\_PHOSPHAT\_1; 1.  
DR PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.  
KW Hydrolase; Periplasmic; Signal.  
FT SIGNAL 1 23 POTENTIAL.  
FT CHAIN 24 417 GLUCOSE-1-PHOSPHATASE.  
FT ACT SITE 42 42 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).  
FT ACT SITE 314 314 PROTON DONOR (BY SIMILARITY).  
SQ SEQUENCE 417 AA; 46940 MW; 8CC4CA4F55DBF90 CRC64;  
  
Query Match 15.5%; Score 256; DB 1; Length 417;  
Best Local Similarity 29.0%; Pred. No. 1.1e-12; Indels 94; Gaps 27;  
Matches 119; Conservative 57; Mismatches 140;  
  
QY 1 VSLLAALLAALAPAAAAE---PLKLEKVLISRHGRVSP--TKALMVS-----PD 49  
DB 6 LTLCLSAAL-PAPIAPTMASTDNQADMYLDQVLVSRHNLRTPIVNTGILTEVTDKQWP 64  
QY 50 WPNVPG-LTPGAALVILGYFARGLLPG-----CPAAGT---TLADVDERT-RTGQ 97  
DB 65 WDAKSGYLTQGGALEVYMGVFWIDQNKLLADELCPTSNEDIYLTNSLQRTIATAQ 124  
QY 98 AFAAGLAPGCAI-VH-----AGDDPIFHGLDT-GCLQDAAILAAGEG---GLTARHGL- 146  
DB 125 FFAAGAFGCKVNIHQPEIGCKDPVFNPIITNSPBEKQKALAAADDYLGSLKXAGE 184  
QY 147 TLAKVLNFA-----SACLECARVVRGLQPL-LASTLSIEIFLEVAQG 187  
DB 185 ELDTVLTAKDSQCKTKDKLCLNLDSCQKNSFIIEADKEFGVSGPLKIANSAVDAILQYEG 244  
QY 188 --PMEVGHGRISA-EM-----LLHNAQFLMRTPIAR---ATPIL-----IVTALS 228  
DB 245 FPAQVANGLVDTTEKWKLNLTAKNAQSTLTFTKIAKNVAHPIILNYIDKGFVSDVKG 304  
QY 229 PARVLLAGHTNLA--LGGDL--WQLPQP-DPTPGGALVFLMN-----RYRVV 275  
DB 305 TAKTIFLVGHDSNIASLMSADFPYQLAQOQYHTPIGKLVFORWTDKQTKDFMKVEY 364  
QY 276 MYQTMQDLR-----NLE-PLPLPRILPIPGC--GSEAACSLSDFARLVAPA 317  
DB 365 VIQTADQRDNAYLSLETPPKHVTLKDCPVDKNGKCSWEDFKVWAKA 414  
  
RESULT 5  
ID PPAL RAT STANDARD; PRT; 423 AA.  
AC P20611.  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE Lysosomal acid phosphatase precursor (EC 3.1.3.2) (LAP).  
GN AC2.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89350910; PubMed=2764916;  
RA Himeno M., Fujita H., Noguchi Y., Kono A., Kato K.;  
RT "Isolation and sequencing of a cDNA clone encoding acid phosphatase  
RL in rat liver lysosomes."  
RL Biochem. Biophys. Res. Commun. 162:1044-1053(1989).  
CC -!- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an  
CC alcohol + phosphate.  
CC -!- SUBCELLULAR LOCATION: Lysosomal.  
CC -!- SIMILARITY: Belongs to the histidine acid phosphatase family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M27893; AAA40744.1; -  
CC PIR; A33395; A33395.  
CC HSP; P20646; IRPA.  
CC InterPro; IPR000560; HisAc\_phosphatase.  
CC Pfam; PF00328; acid\_phosphat; 1.  
CC PROSITE; PS00616; HIS\_ACID\_PHOSPHAT\_1; 1.  
CC PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.  
KW Hydrolase; Signal; Glycoprotein; Lysosome.  
FT SIGNAL 1 30  
FT CHAIN 31 423 LYSOSOMAL ACID PHOSPHATASE.  
FT DISULFID 159 370 BY SIMILARITY.  
FT DISULFID 212 310 BY SIMILARITY.  
FT DISULFID 345 349 BY SIMILARITY.  
FT ACT SITE 42 42 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).  
FT ACT SITE 84 84 BY SIMILARITY.  
FT ACT SITE 286 286 PROTON DONOR (BY SIMILARITY).  
FT CARBOHYD 92 92 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 133 133 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 167 167 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 177 177 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 191 191 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 197 197 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 267 267 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 322 322 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 331 331 N-LINKED (GLCNAC. .) (POTENTIAL).  
SQ SEQUENCE 423 AA; 48319 MW; 6F4C4819CADD4496 CRC64;  
  
Query Match 8.2%; Score 134.5; DB 1; Length 423;  
Best Local Similarity 25.3%; Pred. No. 0.0029;  
Matches 95; Conservative 33; Mismatches 157; Indels 91; Gaps 20;  
  
QY 4 LLAALLAALLAALAPAAAAEPLKLEKVLISRHGRVSPKALMVSP--DWPVPG---LTP 58  
DB 13 LLQFLGMLVMVPIQARSLRF--VTLRYHGDSRPVKAYPKDPYQEKWPGQGLTK 70  
QY 59 RGAAL-----VLGVRYFARGLLPGCPAAGTTL---ADVDERTRTGOAPAGLAPCAIV 110  
DB 71 EGMQLHWELGQALRQRY--HGFNLSYHROVYVSTDFDRLMSABANLAGLFPFTEVQ 128  
QY 111 H-----AGDDPIFHGLDTG-----CLDQADAI-L 133  
DB 129 HFNPNISWQPIPVHTVPTEDRLKFPGLGCPRYEQQLQNETRQTEPYQNMISQNAQFLDM 188  
QY 134 AAGEGLTARHGLTLAKVLNFAACLECARVVRGLGPLAS-----TLSEI-----FLLEY 184  
DB 189 VANETGLM---NLLETIMNVDT-LFCEQTHGLLLPWPASQTVQALSQDKDFSLFLP 244  
QY 185 AQCPMEVGMGRISAELWLLHNAQFLMNRTPYIARATPILIVTALS PARVVLLAGHTNLA 244  
DB 245 GIHD-QVQKARLQGGVLL--AQILKNLT-----LMATTSPQKLLVYSAHDTTLVA 292  
QY 245 GGLDLWQLPQDETTPGGALVELM-----NRYRVVMYQTMQDLRNLPLPRILPIPGCG 300  
DB 293 LQWAL--NVYNGKQAPYASCHIFEOYDNGNGNSVEMY-----FRNDSKAPMPLTLPGCP 346  
301 SEAAACSLSDFARLVAP 316  
347 HR--CPQLQDFLRLTEP 360  
  
RESULT 6  
ID PPAL MOUSE STANDARD; PRT; 423 AA.  
AC P24638; Q8QZT5;  
DT 01-MAR-1992 (Rel. 21, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Lysosomal acid phosphatase precursor (EC 3.1.3.2) (LAP).  
GN ACP2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RC SEQUENCE FROM N.A.  
RP MEDLINE=22388257; PubMed=12477932;  
RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S.E., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettner M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [2]  
RP SEQUENCE OF 3-423 FROM N.A.  
RX MEDLINE=91282986; PubMed=2059337;  
RA Geier C., von Figura K., Pohlmann R.;  
RT "Molecular cloning of the mouse lysosomal acid phosphatase.";  
RL Bio. Chem. Hoppe-Seyler 372:301-304 (1991).  
CC -|- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an  
alcohol + phosphate.  
CC -|- SUBCELLULAR LOCATION: Lysosomal.  
CC -|- SIMILARITY: Belongs to the histidine acid phosphatase family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
DR EMBL; BC023343; AAR23343.1; -;  
DR EMBL; X57199; CAA040485.1; -;  
DR F01; S14742; S14742.  
DR HSP; P15309; 2HPA.  
DR MGD; MGI:87882; ACP2.  
DR InterPro: IPR000560; HisAc\_phsphtse.  
DR Pfam; PF00328; acid\_phosphat; 1.  
DR PROSITE; PS00616; HIS\_ACID\_PHOSPHAT\_1; 1.  
DR PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.  
KW Hydroxylase; Signal; Glycoprotein; Lysosome.  
FT SIGNAL 1 30  
FT CHAIN 31 423 LYSOSOMAL ACID PHOSPHATASE.  
FT DISULFID 159 370 BY SIMILARITY.  
FT DISULFID 212 310 BY SIMILARITY.  
FT DISULFID 345 349 BY SIMILARITY.  
FT ACT SITE 42 42 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).  
FT ACT SITE 84 84 BY SIMILARITY.  
FT ACT SITE 286 286 PROTON DONOR (BY SIMILARITY).  
FT CARBOHYD 92 92 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 133 133 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 167 167 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 177 177 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 191 191 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 197 197 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 267 267 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 322 322 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 331 331 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 374 374 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CONFLICT 131 131 N -> S (IN REF. 2).  
SQ SEQUENCE 423 AA; 48508 MW; AFD7ASC90A4F2AF4 CRC64;  
Query Match 7.7%; Score 127.5; DB 1; Length 423;  
Best Local Similarity 23.5%; Pred. No. 0.01;  
Matches 92; Conservative 32; Mismatches 145; Indels 123; Gaps 19;  
Qy 6 AALL-----ALAAPAAARAEPLKLVILSRHVSPTKALMVSP-----DWP----- 51  
Db 11 AALLOFLGCLTVMPIQARS-----LRFVTLRYHGRSPVKTYPKDQYQEKMPQGF 66  
Qy 52 -----WPVGLTPRGAALVLLGYR-----FARGLLPGCPAAGTTLADVDERTRG 96  
Db 67 QLTKEGMLQHWELGQALRQRYHGFNTSVHRQEVYRS-----TDFDRLMSA 114  
Qy 97 QAFAGLAPGCAIVH-----AGDDPIFHL----- 121  
Db 115 EANLAGLFPNEVQHFNPNISWQIPVHTVPTEDRLKPLGCPRYFOLQNETQTPE 174  
Qy 122 -DTGCLDQADAI-LAAGEGLTARHGLTLAKVLFASFACLECARVVRGLPLIAS----- 174  
Db 175 YQNRSTQNAQFLNVAETGLT---NVLTETIWNVYDT-LFCEQTHGLLLPPWASPTVQ 230  
Qy 175 -----TLSEIFLEVAQSPMEVWGRIASAEWLLHNAQFLMRTPIARATPILIVTALS 228  
Db 231 RLSQLKDFSLFLFGIHE---QVKARLQGGVLL---AQILKNLT-----LMATTSQ 276  
Qy 229 PARVILLAGHTNLALGGLDLWOLPQDPTTPPGALVFLW-----NRYVVMYQTMDQLR 284  
Db 277 FPKLLVYSAHDTTLVALQMAL-NVYNGKQAPYASCHIFELYQEDNGNFSVMY-----FR 330  
Qy 285 NLSEFLPRLPIPGCGSEACSLSDPARLVAP 316  
Db 331 NDSKQAPWPLILPGCPHR--CPLQDFLRLTEP 360  
RESULT 7  
PPAL\_HUMAN STANDARD; PRT; 423 AA.  
ID AC P1117; Q9BTU7;  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Lysosomal acid phosphatase precursor (EC 3.1.3.2) (LAP).  
GN ACP2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RC SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE=89052645; PubMed=3191910;  
RA Pohlmann R., Krentler C., Schmidt B., Schroeder W., Lorkowski G.,  
RA Culley J., Merzmann G., Geier C., Waheed A., Gottschalk S.,  
RA Grzeschik K.H., Hasikik A., von Figura K.;  
RT "Human lysosomal acid phosphatase: cloning, expression and  
chromosomal assignment.";  
RL EMBL J. 7:2343-2350 (1988).  
RN [2]  
RC SEQUENCE FROM N.A.  
RX MEDLINE=93377828; PubMed=2776754;  
RT Geier C., von Figura K., Pohlmann R.;  
RL "Structure of the human lysosomal acid phosphatase gene.";  
RN Eur. J. Biochem. 183:611-616 (1989).  
RP SEQUENCE FROM N.A.  
RC TISSUE=Kidney;

RX MEDLINE=22388257; PubMed=12477932;  
RA Krausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shennen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Besak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Vallalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Whiting J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Blakesley R.M., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S., Krzywinski M.I., Skalska U., Smalley D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RL and mouse cDNA sequences";  
CC -1- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an  
CC alcohol + phosphate.  
CC -1- SUBCELLULAR LOCATION: Lysosomal.  
CC -1- SIMILARITY: Belongs to the histidine acid phosphatase family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
DR EMBL; X12548; CAA31064.1; --  
DR EMBL; X15525; CAA33542.1; JOINED.  
DR EMBL; X15526; CAA33542.1; JOINED.  
DR EMBL; X15527; CAA33542.1; JOINED.  
DR EMBL; X15528; CAA33542.1; JOINED.  
DR EMBL; X15529; CAA33542.1; JOINED.  
DR EMBL; X15530; CAA33542.1; JOINED.  
DR EMBL; X15531; CAA33542.1; JOINED.  
DR EMBL; X15532; CAA33542.1; JOINED.  
DR EMBL; X15533; CAA33542.1; JOINED.  
DR EMBL; X15534; CAA33542.1; JOINED.  
DR EMBL; X15535; CAA33542.1; JOINED.  
DR EMBL; BC003160; AA03160.1; --  
DR PIR; S06167; S06167.  
DR HSP; P20646; IRPA.  
DR Genew; HGNC:123; ACP2.  
DR MW; J71650; --  
DR GO; GO:0016021; C: integral to membrane; TAS.  
DR GO; GO:0003993; F: acid phosphatase activity; TAS.  
DR InterPro; IPR000560; HisAcPhosphatase.  
DR Pfam; PF00328; acid\_phosphat; 1.  
DR PROSITE; PS00616; HTS\_ACID\_PHOSPHAT\_1; 1.  
DR PROSITE; PS00778; HTS\_ACID\_PHOSPHAT\_2; 1.  
KW Hydrolase; Signal; Glycoprotein; Lysosome.  
FT SIGNAL 1 30  
FT CHAIN 31 423  
FT DISULFID 159 370  
FT DISULFID 212 310  
FT DISULFID 345 349  
FT ACT\_SITE 42 42  
FT ACT\_SITE 84 84  
FT ACT\_SITE 286 286  
FT ACT\_SITE 92 92  
FT CARBOHYD 133 133  
FT CARBOHYD 167 167  
FT CARBOHYD 177 177  
FT CARBOHYD 191 191  
FT CARBOHYD 267 267  
FT CARBOHYD 322 322

FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 29 29 Q -> R (IN REF. 1 AND 2).  
SQ SEQUENCE 423 AA; 48316 MW; 1A2CBFB40ED3724B CRC64;  
Query Match 7.6%; Score 124.5; DB 1; Length 423;  
Best Local Similarity 23.9%; Pred. No. 0.017;  
Matches 95; Conservative 32; Mismatches 136; Indels 135; Gaps 23;  
QY 4 LIAALLA--LAALAPAAAAEPLKELKRVLSRHGVRSPYKALMVSP-----DWP----- 51  
DB 13 LQLLLGVNLVMPPTQARS-----LRFVTLRYHGRDSPVKTYPKDPYQBEENPQFGQOL 68  
QY 52 -----WPVGLTPRGAALVLLGVY-----FARGLLPGCPAAGTILADVDRTRTGOA 98  
DB 69 TKEGMLQHWELGQALRQRYHGFNLTSYHQEVYRS-----TDFRILMSAEA 116  
QY 99 FAAGLAPGCAI-----VHA-----GDDPIFHGL----- 121  
DB 117 NLAGLPPNGQRFNPINSMQPIFVHTVPTEDRLKFLPGLPCRYEQLOQNETQTPEYQ 176  
QY 122 -----DTGCLDQADAILAAGEGGLTARHGLTLAKVLFASFACLCARVVG-RLGPLLA-S 174  
DB 177 NESRRNAQFLD-----MVANETGLT-----DLFETVWVNYDT-LFCQTHGLRLPPWASPQ 227  
QY 175 TLSIFLLE-----YAQGPMEVGMGRISAELLHNAQFLMNRTPFYARATPILIV 224  
DB 228 TMQRLSLRLKDFSRFLFGIYQAEK-----ARLOGGVLL--AQIRKULT-----LMA 272  
QY 225 TALSPARVVLVLAGHDTNLAL--GGDLWLQLPQDPETPPGGALVELW-----NRVYRVYQ 278  
DB 273 TTSQFLKLLVYSADHTTUVALQMALDVI-----NGSQAPVYASCHIFELIQEDSGNFSVMY- 328  
QY 279 TMDQRLNLEPLLRILPIPIFGGSEAAACSLSDFAFLVAP 316  
DB 329 ----FRNESDKAPWPLSLPGCFHR--CPLODFLRLTEP 360  
RESULT 8  
CTPD\_MYCTU STANDARD; PRT; 657 AA.  
ID CTPD\_MYCTU  
AC O53160;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Probable cation-transporting 2-type ATPase D (EC 3.6.3.-).  
GN CTPD OR RV1469 OR MT1515 OR MT0007.16 OR MB1504.  
OS Mycobacterium tuberculosis, and  
OS Mycobacterium bovis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773, 1765;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=M.tuberculosis; STRAIN=H37Rv;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eigemeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies K., Devlin K., Feldwell T., Gentles S., Hamlin N., Hoiroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulston J.B., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence.";  
RL Nature 393:537-544 (1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;  
RX MEDLINE=22206494; PubMed=12218036;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,

Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A., Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.W.; "Whole-genome comparison of Mycobacterium tuberculosis clinical and laboratory strains."; J. Bacteriol. 184:5479-5490(2002).

[3]

SEQUENCE FROM N.A.

RP SPECIES-M.bovis; STRAIN-AF2122/97;

RC MEDLINE=22709107; PubMed=12788972;

RA Garnier T., Eglmeier K., Camus J.-C., Medina N., Mansoor H., Pryor M., Duthoy S., Grondin S., Lacroix C., Monsenpe C., Simon S., Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R., Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.; "The complete genome sequence of Mycobacterium bovis."; Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).

RL C-!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.

CC C-!- SUBCELLULAR LOCATION: Integral membrane protein.

CC C-!- SIMILARITY: Belongs to the cation transport ATPases family (P-type ATPases). Subfamily 1B.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

CC EMBL; AL021184; CAA15997.1; -

DR EMBL; AE007021; AAK45780.1; -

DR EMBL; BX248339; CAD96171.1; -

DR PIR; H70872; H70872.

DR TIGR; M1515; -

DR Tuberculist; Rv1469; -

DR InterPro; IPR006416; ATPase-IB\_hvy.

DR InterPro; IPR001757; ATPase\_E1-E2.

DR InterPro; IPR001366; Cad\_ATPase.

DR InterPro; IPR008250; E1-E2\_ATPase\_reg.

DR InterPro; IPR006404; Heavy\_met\_ATPase.

DR InterPro; IPR005834; Hydrolase.

DR Pfam; PF00122; E1-E2\_ATPase; 1.

DR Pfam; PF00702; Hydrolase; 1.

DR PRINTS; PR00119; CATATPASE.

DR PRINTS; PR00941; CDATPASE.

DR TIGRFAMS; TIGR01512; ATPase-IB2\_Cd; 1.

DR TIGRFAMS; TIGR01525; ATPase-IB\_hvy; 1.

DR TIGRFAMS; TIGR01494; ATPase\_P-type; 2.

DR PROSITE; PS00154; ATPASE\_E1\_E2; 1.

DR Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding; Complete proteome.

KW TRANSNEM 40 60 POTENTIAL.

FT TRANSNEM 62 82 POTENTIAL.

FT TRANSNEM 102 122 POTENTIAL.

FT TRANSNEM 268 288 POTENTIAL.

FT TRANSNEM 299 319 POTENTIAL.

FT TRANSNEM 447 467 POTENTIAL.

FT TRANSNEM 477 497 POTENTIAL.

FT TRANSNEM 607 627 POTENTIAL.

FT MOD\_RES 347 347 PHOSPHORYLATION (BY SIMILARITY).

FT METAL 543 543 MAGNESIUM (BY SIMILARITY).

FT METAL 547 547 MAGNESIUM (BY SIMILARITY).

SC SEQUENCE 657 AA; 67984 MW; 1AF5A7DD4BC697D5 CRC64;

Query Match 6.9%; Score 113.5; DB 1; Length 657;

Best Local Similarity 20.4%; Pred. No. 0.19;

Matches 80; Conservative 50; Mismatches 125; Indels 137; Gaps 17;

QY 8 LLALALAPAAAAEPLKLEKVVLT-----SRHGVR-----PTKALMVP 48

Db 96 LLMIAAVGAIVAIQIDFGALLIVIFATSGALDDIATRHATSAESVKGLLDIAPDQAVVQ 155

QY 49 DWHPVGLTPRGALVLLGYYRFARGLLPG---CPAAGTIL---ADVDERTETGQAFAGL 103





```
QY 99 FAAGLAP--CAAI-----VHA---GDD-----PIFHGLDTGCLDQ----- 128
Db 119 NLAAFPPEGVSWNPILLQPIVHTVPLSEQLLYLPFRNCPFELESETLKSBEFQ 178
QY 129 -----ADAILAAGEGLTARHGLTILAKVLNFPASACLECARVVGRLGPLLAS----- 174
Db 179 KELHPYKDFATLGK--LSGLHQDLFGIWSKYVDPLYCESVHNFTLPSWATEDMTKLR 236
QY 175 TLSEIFLELYAQQPMVGVGRISAEMLLHNAQFLMRTPIYIARATPILIVTALSPPARVL 234
Db 237 ELSELSLSLYGHKHQKESRLOGGLVNE---ILN---HMKRATQI-----PSYKKLIM 285
QY 235 LAGHTNIALGGLDLWQLPQDPTPPGCA-----LVFELMNRVVRVYQTMQDLRLNLEPL 289
Db 286 YSAHDTTVS--GLQVLDVYNGLLPPVASCHELYFERGEYFVEMY-----RNETQH 337
QY 290 LPRILPIPGCGSBAACSLSPFARLVAP 316
Db 338 EYPPLMLPGC--SPSCPLERFAELVGP 362

RESULT 10
FMT_BORBR
ID FMT_BORBR STANDARD; PRT; 312 AA.
AC Q7WQSS;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DE Methionyl-tRNA formyltransferase (EC 2.1.2.9).
GN FMT OR B50248.
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RB50 / ATCC BAA-588;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.M., Temple L., James K., Bason N., Cherevach I.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Leather S., Moutle A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Rabinowitsch E., Rutter S., Saunders M., Saunders R., Squares S., Stevens K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RA "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
CC -!- FUNCTION: Modify the free amino group of the aminoacyl moiety of
CC methionyl-tRNA(fMet). The formyl group appears to play a dual role
CC in the initiator identity of N-formylmethionyl-tRNA by: (i)
CC promoting its recognition by IF2 and (ii) impairing its binding to
CC EFTu-GTP (By similarity).
CC -!- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + L-methionyl-
CC tRNA(fMet) + H(2)O = tetrahydrofolate + N-formylmethionyl-
CC tRNA(fMet).
CC -!- SIMILARITY: Belongs to the fnt family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC ENBL; BX640437; CAB30746.1; -.
CC HAMAP; MF 00182; -.
CC InterPro; IPR005794; Fmt.
```

```
DR InterPro; IPR002376; formyl_transf.
DR InterPro; IPR005793; formyl_transf_C.
DR Pfam; PF00551; formyl_transf; 1.
DR Pfam; PF02911; formyl_transf_C; 1.
DR TIGRFAMs; TIGR00460; fnt; 1.
DR PROSITE; PS00373; GART; FALSE NEG.
KW Transferase; Methyltransferase; Protein biosynthesis;
KW Complete proteome.
FT BINDING 117..120 TETRAHYDROFOLATE (THF) (BY SIMILARITY).
SQ SEQUENCE 312 AA; 32601 MW; 1F70614C084D7B55 CRC64;

Query Match 6.2%; Score 102.5; DB 1; Length 312;
Best Local Similarity 25.4%; Pred. No. 0.64;
Matches 52; Conservative 24; Mismatches 66; Indels 63; Gaps 10;

QY 9 LALAALAPAAARAEPLKLEK-----VWLSRHGVSPKALMVSPDWMPV----- 54
Db 65 LRLDGRYPDEAAARAQALRVAPDMVMVYAG-----LILQWTLDFRLGLNIHA 116
QY 55 GLTP--RGAALVILGYRYFARGLLPGCPAAGTILADVDERTTG-----QAFAGLAPGC 107
Db 117 SLLPWRGAAP-----QRAIAGDAETGVTIMQMDAGLDTGDMLERAVPICAOQA 169
QY 108 AIVH-----AGDDPIFHGLDTGCLDQADAILAAGEGGITARH-----GLTLAKVLNFSAC 158
Db 170 AOLHDELALAGGQAI-----VDALAAGQGGGLAPRRQDPDAGVTTAAKLDKAA 218
QY 159 LEC-----ARVVGRLGPLLAATL 176
Db 219 LDCSLPRAAVLARRVRAFNVPVGAT 243

RESULT 11
FMT_BORPA
ID FMT_BORPA STANDARD; PRT; 312 AA.
AC Q7MIV2;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Methionyl-tRNA formyltransferase (EC 2.1.2.9).
GN FMT OR BPF0244.
OS Bordetella parapertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=519;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=12822 / ATCC BAA-587;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.M., Temple L., James K., Bason N., Cherevach I.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Fellwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leather S., Moutle S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Saunders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RA "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
CC -!- FUNCTION: Modify the free amino group of the aminoacyl moiety of
CC methionyl-tRNA(fMet). The formyl group appears to play a dual role
CC in the initiator identity of N-formylmethionyl-tRNA by: (i)
CC promoting its recognition by IF2 and (ii) impairing its binding to
CC EFTu-GTP (By similarity).
CC -!- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + L-methionyl-
CC tRNA(fMet) + H(2)O = tetrahydrofolate + N-formylmethionyl-
CC tRNA(fMet).
CC -!- SIMILARITY: Belongs to the fnt family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```



CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL; BX640423; CAE39985.1; AUT\_INIT.  
 CC

DR HAMAP; MF 00182; -; 1.  
 DR InterPro; IPR005794; Fmt.  
 DR InterPro; IPR005376; formyl\_transf.  
 DR InterPro; IPR005793; formyl\_transf. C.  
 DR Pfam; PF00551; formyl\_transf. 1.  
 DR Pfam; PF02911; formyl\_transf. C. 1.  
 DR TIGRFAMs; TIGR00460; fnt; 1.  
 DR PROSITE; PS00373; GART; FALSE NEG.

KW Transferase; Methyltransferase; Protein biosynthesis;  
 KW Complete proteome.

FT BINDING 117 TETRAHYDROPOLATE (THP) (BY SIMILARITY).  
 SQ SEQUENCE 312 AA; 32529 MW; 1F7060A6084D7BDF CRC64;

Query Match 6.2%; Score 102.5; DB 1; Length 312;  
 Best Local Similarity 25.4%; Pred. No. 0.64;  
 Matches 52; Conservative 24; Mismatches 56; Indels 53; Gaps 10;

QY 9 LALAALAPAAAAPLKEK-----VILSRHGVSPTKALMVSPDWPV----- 54  
 DB 65 LRLDGRYPDEAAAAAQLERVAPDMVVAAYG-----LILPQWTLPLRLGCLNIHA 116

QY 55 GLTP--RGAALLGLYRFGALPLGCPAAGTILADVDRTRTG-----QAFAGLAPGC 107  
 DB 117 SLLEPRWGAIFI-----QRAIAGDAETGVITQMDAGLDTGMDLERAVPIGAQTA 169

QY 108 AIVH-----AGDDPIFHGLDTGCLDQADAILAAGEGLTARH-----GLTLAKVLNFAAC 158  
 DB 170 AQLHDELALAGQAI-----VDALAALGGGLAPRRQPDAGVTVAALKDKAEA 218

QY 159 LEC-----ARVVGRLGPLLSTL 176  
 DB 219 LDCSLPAAVLARRVRAFNPFVGAT 243

RESULT 12  
 YOH2 STRCO STANDARD; PRT; 411 AA.

AC P40180; Q9S6U5;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Putative glycosyl transferase SC03672 (EC 2.-.-.-).  
 GN SC03672 OR SCH35.52.

OS Streptomyces coelicolor.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycinae; Streptomycetaceae; Streptomyces.

OX NCBI\_TaxID:1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 EX MEDLINE=96293864; PubMed=8722574;  
 RA Brans A., Loriaux A., Joris B., Dusart J.;

RT "Cloning and sequencing of the dhak locus in Streptomyces coelicolor  
 AT A3(2).";  
 RL DNA Seq. 6:179-184(1996).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 EX MEDLINE=21996410; PubMed=12000953;

RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,  
 RA Rabinowitch E., Rajandream M.A., Rutherford K., Rutter S.,

RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorrek A., Woodward J., Barrall B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 coelicolor A3(2).";  
 RL Nature 417:141-147(2002).

CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -1- SIMILARITY: Belongs to the glycosyltransferase family 4.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL; X77458; CAAS4605.1; -;  
 CC EMBL; AL939117; CAB44375.1; -;  
 CC FIR; S41945; S41945.

DR InterPro; IPR00715; Glyco trans 4.  
 DR Pfam; PF00953; Glycos transf 4; 1.

KW Hypothetical protein; Transferase; Transmembrane; Complete proteome.

FT TRANSMEM 7 27 POTENTIAL.  
 FT TRANSMEM 45 65 POTENTIAL.

FT TRANSMEM 70 90 POTENTIAL.  
 FT TRANSMEM 120 140 POTENTIAL.

FT TRANSMEM 148 168 POTENTIAL.  
 FT TRANSMEM 169 189 POTENTIAL.

FT TRANSMEM 197 217 POTENTIAL.  
 FT TRANSMEM 227 247 POTENTIAL.

FT TRANSMEM 277 297 POTENTIAL.  
 FT TRANSMEM 301 321 POTENTIAL.

FT CONFLICT 195 195 A -> R (IN REF. 1).  
 SQ SEQUENCE 411 AA; 41127 MW; A013111F7BC1CA9B CRC64;

Query Match 6.1%; Score 100.5; DB 1; Length 411;  
 Best Local Similarity 25.5%; Pred. No. 1.2;  
 Matches 74; Conservative 31; Mismatches 98; Indels 87; Gaps 15;

QY 2 SLLAALALALAPAAAAPLKEKVVILSRHGVSPTKALMVSPDWPVGLTPRGA 61  
 DB 17 SFLLTAVLAALLRAPAL-----RLAVDRRRRR-----EVPLL-GGV 52

QY 62 ALVLL-----GYRFGALPLGCPAAGTII-----ADVDETRTQGAFAAG 102  
 DB 53 AVLVITAVVANAGDR---AGVPLGPAGRLLAATVVGALGLAADVWELRRLLAGTA 109

QY 103 LAPGCAIVHAGDDPIFHGLDTG-----LPQADAILAAGEGLTARHGLTLAKYL 152  
 DB 110 VAAACVVPYGETGPGVAGALAVGWALVTGAFRGLDHADGW--GTVGVVTAFGVG----- 162

QY 153 NFASACLECARVVGRLGP-----LLASTLSEIFLLEYAQGPMEVG-WGEISAEWLLHNAQ 206  
 DB 163 -----ACAAVELMDGPAVLLVLAALAGFLHNHHPARIALGACGLFTGLTGA 215

QY 207 FLNMRTPY--IARATPILITVLTALSPARVVL-----LAGHDTNLALGLLD 248  
 DB 216 VLA-RTGYGPGVGAGVLCALTAVPVADAIVLLSRLLAGRP--LSRGGPD 262

RESULT 13  
 MOT3 RAT STANDARD; PRT; 492 AA.

AC 070461;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Monocarboxylate transporter 3 (MCT 3).  
 GN SLC16A8 OR MCT3.

OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Retina;  
 RX MEDLINE=99020931; PubMed=9841555;  
 RA Philip N.J., Yoon H., Grollman E.F.;  
 RT "Monocarboxylate transporter MCT1 is located in the apical membrane  
 of MCT3 in the basal membrane of rat RPE.";  
 RL Am. J. Physiol. 274:R1824-R1828 (1998).  
 CC -|- FUNCTION: Proton-linked monocarboxylate transporter. Catalyzes the  
 rapid transport across the plasma membrane of many  
 monocarboxylates such as lactate, pyruvate, branched-chain oxo  
 acids derived from leucine, valine and isoleucine, and the ketone  
 bodies acetoacetate, beta-hydroxybutyrate and acetate (By  
 similarity).  
 CC -|- SUBCELLULAR LOCATION: Integral membrane protein. Plasma membrane.  
 CC -|- TISSUE SPECIFICITY: RETINAL PIGMENT EPITHELIUM. RESTRICTED TO THE  
 BASOLATERAL MEMBRANE.  
 CC -|- SIMILARITY: Belongs to the major facilitator superfamily.  
 CC Monocarboxylate porter (TC 2.A.1.13) family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF059258; AAC18120.1; --  
 DR InterPro; IPR007114; MFS.  
 DR InterPro; IPR004743; Monocarb transp.  
 DR TIGRFAMs; TIGR00892; 2A0113; 1.  
 DR PROSITE; PS0850; MFS; 1.  
 KW Transport; Symport; Transmembrane; Multigene family.  
 FT DOMAIN 1 14 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 15 35 POTENTIAL.  
 FT DOMAIN 36 58 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 59 79 POTENTIAL.  
 FT DOMAIN 80 85 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 86 106 POTENTIAL.  
 FT DOMAIN 107 115 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 116 136 POTENTIAL.  
 FT DOMAIN 137 147 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 148 168 POTENTIAL.  
 FT DOMAIN 169 172 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 173 193 POTENTIAL.  
 FT DOMAIN 194 230 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 231 251 POTENTIAL.  
 FT DOMAIN 252 257 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 258 278 POTENTIAL.  
 FT DOMAIN 279 293 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 294 314 POTENTIAL.  
 FT DOMAIN 315 318 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 319 339 POTENTIAL.  
 FT DOMAIN 340 352 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 353 373 POTENTIAL.  
 FT DOMAIN 374 386 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 387 407 POTENTIAL.  
 FT DOMAIN 408 492 CYTOPLASMIC (POTENTIAL).  
 FT SEQUENCE 492 AA; 51590 MW; 5CCC4DC3835FA0 CRC64;

Query Match 6.0%; Score 98.5; DB 1; Length 492;  
 Best Local Similarity 25.5%; Pred. No. 2.1;  
 Matches 79; Conservative 29; Mismatches 123; Indels 79; Gaps 16;

OY 48 PD--WNPVGLTPRGAALVLLGVYRPARGLPGCP--AAGTILADVTRTGTGAFAAGLA 104  
 DB 13 PDGGWGVV-----LGACFVITGAY-----GFPKAVSVFFRELKEDFGAGYSDTAWVS 61  
 OY 105 PGCAIVHAGDDPIFGHLDLT--GCLDQADAILAGEGGLTARHGLTAKVINFASACIE-- 160  
 DB 62 SIMLAMYGTGPISSILVTRFGC---RPVMLA---GGLLASAGMILA---SFAIRLLELY 112

OY 161 -CAVVVGRGLGPIASTLSEIFLLEYAGQPMVGMGRISAEWLLHNAQLMNRTPYIARAT 219  
 DB 113 LTAGVLTGLGLALNFQPSLIMGLYFERRPLANGLA--GS 153  
 OY 220 PILIVTALSAPRVL---LAGHDTNIALGGLDL-----WQLPQDPETPPGG- 262  
 DB 154 PVFLST--LSPGLQLGERFGWRGGLFLFGLLHCCACGAVMPPPPQPRPDAPPGGR 212  
 OY 263 -----ALVPFLMWRVYVMYQTMQDLRNLEPLLPRL- -PIPGGSEAAACSLSD 309  
 DB 213 ARURQLDLAVCTDRFMVNTVKFLMALGLFVPAILLVNYAKDAGVDAFAAFLLSIVG 272  
 OY 310 FARLVA-PAC 318  
 DB 273 FVDIVARPAAC 282

## RESULT 14

KAS2\_STRCM ID KAS2\_STRCM STANDARD; PRT; 402 AA.  
 AC P41176;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Putative polyketide beta-ketoacyl synthase 2 (EC 2.3.1.-) (ORF 2).  
 OS Streptomyces cinnamonensis.  
 CC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 CC Streptomycineae; Streptomycetaceae; Streptomyces.  
 CC NCBI\_TaxID=1900;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3823.5;  
 RX MEDLINE=92374994; PubMed=1508151;  
 RA Arrowsmith T.J., Malpartida F., Sherman D.H., Birch A., Hopwood D.A.,  
 RT Robinson J.A.;  
 RT "Characterisation of acti-homologous DNA encoding polyketide synthase  
 genes from the monensin producer Streptomyces cinnamonensis.";  
 RL Mol. Gen. Genet. 234:254-264 (1992).  
 CC -|- PATHWAY: Polyketide antibiotic monensin biosynthesis.  
 CC -|- MISCELLANEOUS: This putative ketoacyl synthase lacks the active  
 site cysteine.  
 CC -|- SIMILARITY: Belongs to the beta-ketoacyl-ACP synthases family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; Z11511; CAA77597.1; --  
 DR PIR; S25077; S25077.  
 DR HSRP; P73283; 185M.  
 DR InterPro; IPR000794; Ketoacyl\_synth.  
 DR Pfam; PF00109; Ketoacyl-synt; 1.  
 DR Pfam; PF02801; Ketoacyl-synt; 1.  
 KW Antibiotic biosynthesis; Transferase; Acyltransferase.  
 SQ SEQUENCE 402 AA; 41270 MW; 348657D10201C214 CRC64;

Query Match 6.0%; Score 98; DB 1; Length 402;  
 Best Local Similarity 21.8%; Pred. No. 1.8;  
 Matches 69; Conservative 37; Mismatches 133; Indels 78; Gaps 12;

OY 23 FLKLEKVVILSRHGVRSPTKALMYSPPWPMFVGLTPRGAALVLL-----GYRPFARGLLP 77  
 DB 3 PVAVTGMGIAAPNGLGRPTTG---RPPW-----APRAASAASRTRFDPSGYPQALAGEIP 53  
 OY 78 GCPRA-----GTLADVTRTGTGAFAAGLAPGCAIVHAGDDPIFGHLDTGCLDQADAIL 133  
 DB 54 GFRAAEHLPGRLVPTQDRVTRLSLAADWALADAGVEVAEDP-----LDWG-----VVT 103

134	QY	QY	AAGGEGLTARHGLTLAKVLNFPASACLECARVWGRLGPELLASTSEIFLLEYAQGPMEVCW	193
			:	:
104	Dbb	Dbb	ASHAGG-----FEQODELOKLLGGQFVLISAYOSFAMFYAVNSGGIISIRH	149
			:	:
			:	:
194	QY	QY	GRISAEMLLHNAQ-----FLMNRFPYIARATPILITVLTSPARVLLAGHDTNLALGGLD	248
			:	:
			:	:
150	Dbb	Dbb	GKKGPSGVVWSDQAGGLDALAQARLVRKGTPLIVCGAEP-RSAPGASPSPPA-GGMS	207
			:	:
			:	:
249	QY	QY	LWQLPQPDDEP-----PGGALVPELWNRYRVVYQTMDOGLRNLEPLLPRILPIPGC	299
			:	:
208	Dbb	Dbb	DSDEFENRAYLFFDRDGRGYPPGG-----RGVVPLEAEAPAR	247
			:	:
300	QY	QY	GSEAAACSLSDPARLVAP	316
			:	:
			:	:
248	Dbb	Dbb	GAEVYGEAGAPLARLPAP	264

## RESULT 15

```

BC04 ACXY          STANDARD;          PRT; 1307 AA.
Q9WX71;
28-FEB-2003 (rel. 41, Created)
28-FEB-2003 (rel. 41, Last sequence update)
28-FEB-2003 (rel. 41, Last annotation update)
Cellulose synthase 2 operon protein C precursor.
BCSC11.
Acetobacter xylinus.
Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
Acetobacteraceae; Gluconacetobacter.
NCBI_TaxID=28448;
[1]_OX
RN
SEQUENCE FROM N.A.
STRAIN=JCM 7664 / IFO 13693;
MEDLINE=99310341; PubMed=10382968;
Umeda Y., Hirano A., Ishibashi M., Akiyama H., Onizuka T., Ikeuchi M.,
Inoue Y.;
"Cloning of cellulose synthase genes from Acetobacter xylinum JCM
7664: implication of a novel set of cellulose synthase genes.";
DNA Res. 6:109-115(1999).
-!- FUNCTION: Required for maximal bacterial cellulose synthesis (By
similarity).
-!- PATHWAY: Bacterial cellulose biosynthesis.
-!- SUBCELLULAR LOCATION: Outer membrane (Potential).
-!- SIMILARITY: Belongs to the acsc/bcsc family.
-!- SIMILARITY: Contains 9 TPR repeats.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation on its
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial/
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AB015803; BAA77596.1; -.
InterPro; IPR008410; BCSC_C.
InterPro; IPR003921; Cell_synth_C.
InterPro; IPR008941; TPR_Like.
InterPro; IPR001440; TPR.
Pfam; PF05420; BCSC_C; 1.
Pfam; PF00515; TPR; 5.
PRINTS; PR01441; CELLSYNTHASEC.
SMART; SM0028; TPR; 4.
Cellulose biosynthesis; Repeat; TPR repeat; Membrane; Outer membrane;
Signal.
1 55
FT CHAIN 56 1307
FT REPEAT 97 130
FT REPEAT 270 303
FT REPEAT 339 372
FT REPEAT 374 406
FT REPEAT 458 491
FT REPEAT 493 525
FT REPEAT 1 55
FT CHAIN 56 1307
FT REPEAT 97 130
FT REPEAT 270 303
FT REPEAT 339 372
FT REPEAT 374 406
FT REPEAT 458 491
FT REPEAT 493 525
CELLULOSE SYNTHASE 2 OPERON PROTEIN C.

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model  
Run on: April 30, 2004, 12:34:35 ; Search time 28.8094 Seconds  
(without alignments)  
3482.836 Million cell updates/sec

Title: US-10-021-723B-14  
Perfect score: 1647  
Sequence: 1 VSLLAALLAALAPAAAA.....CGSEAACSLSDFARLVAPAC 318

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues  
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database :
- 1: SP archaea:\*
  - 2: SP bacteria:\*
  - 3: SP fungi:\*
  - 4: SP human:\*
  - 5: SP invertebrate:\*
  - 6: SP mammal:\*
  - 7: SP nhc:\*
  - 8: SP organelle:\*
  - 9: SP phage:\*
  - 10: SP plant:\*
  - 11: SP rodent:\*
  - 12: SP virus:\*
  - 13: SP vertebrate:\*
  - 14: SP unclassified:\*
  - 15: SP\_rvirus:\*
  - 16: SP\_bacteriap:\*
  - 17: SP\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	613.5	37.2	432	Q8KD7	Q8kd7 escherichia
2	603.5	37.0	432	Q8KN8	Q8kn8 escherichia
3	607.5	36.9	432	Q8KDB	Q8kdb escherichia
4	607.5	36.9	432	Q8KDB	Q8kdb escherichia
5	607.5	36.9	432	Q8KDB	Q8kdb escherichia
6	606.5	36.8	432	Q8KDB	Q8kdb escherichia
7	606.5	36.8	434	Q8KDB	Q8kdb escherichia
8	605.5	36.8	446	Q8KDB	Q8kdb escherichia
9	603.5	36.6	432	Q8KDB	Q8kdb escherichia
10	600.5	36.5	442	Q8KDB	Q8kdb escherichia
11	596.5	36.2	432	Q8KDB	Q8kdb escherichia
12	590	35.8	441	Q8KDB	Q8kdb escherichia
13	377.5	22.9	414	Q8KDB	Q8kdb escherichia
14	359	21.8	443	Q8KDB	Q8kdb escherichia
15	353	21.4	435	Q8KDB	Q8kdb escherichia
16	348	21.1	392	Q8KDB	Q8kdb escherichia

Q8ppx6 xanthomonas  
Q7dd02 shigella fl  
Q8irv6 shigella fl  
Q8xbz6 escherichia  
Q8z7p1 salmonella  
Q84cn9 klebsiella  
Q8gd20 pseudomonas  
Q7way1 klebsiella  
Q8cw72 escherichia  
Q8vq82 klebsiella  
Q8bz92 homo sapien  
Q8vz93 homo sapien  
Q8vd68 drosophila  
Q81ld5 pseudomonas  
Q8km35 myxococcus  
Q8kuf5 actinosyne  
Q8my06 drosophila  
Q8n015 macaca fasc  
Q8a997 caulobacter  
Q8f3e4 streptomyce  
Q8apf7 legionella  
Q82kh3 rhizobium m  
Q8xwz4 ralstonia s  
Q8u5j7 agrobacteri  
Q8n88 streptomyce  
Q8gk9 homo sapien  
Q8ybp4 aeropyrum p  
Q8ewal streptomyce

ALIGNMENTS

RESULT 1

Q8KD7 PRELIMINARY; PRT; 432 AA.  
ID Q8KD7  
AC Q8KD7  
DT 01-JUN-2002 (TREMBLrel. 21, Created)  
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
DE Periplasmic phosphoanhydride phosphohydrolase.  
GN APPA.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP MEDLINE-90368616; PubMed=2168385;  
RX MEDLINE-90368616; PubMed=2168385;  
RA Dassa J., Marck C., Boquet P.-L.L.;  
RT "The complete nucleotide sequence of the Escherichia coli gene appA  
RT reveals significant homology between pH 2.5 acid phosphatase and  
RT glucose-1-phosphatase.";  
RL J. Bacteriol. 172:5497-5500 (1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-93054596; PubMed=1429631;  
RA Ostanin K., Hams E.H., Stevis P.E., Kuciel R., Zhou M.M.,  
RA Van Etten R.L.;  
RT "Overexpression, site-directed mutagenesis, and mechanism of  
RT Escherichia coli acid phosphatase.";  
J. Biol. Chem. 267:22830-22836 (1992).  
RL EMBL; L03373; AAA0005.1;  
DR GO; GO:0003993; P:acid phosphatase activity; IEA.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR000560; HisAc\_phsphtase.  
DR PFam; PF00328; acid\_phosphat; 1.  
DR PROSITE; PS00616; HIS\_ACID\_PHOSPHAT\_1; 1.  
DR PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.  
DR Hydrolase.  
SQ SEQUENCE. 432 AA; 46971 MW; 755D5E4B1AD916A6 CRC64;

Query Match 37.2%; Score 613.5; DB 2; Length 432;  
Best Local Similarity 45.3%; Pred. No. 9.3e-39;  
Matches 194; Conservative 44; Mismatches 79; Indels 111; Gaps 36;

QY 2 SLLLAALALALAP-AAAAAP-LKLEKVVILSRHGRVSPTKA--LM--VSPD-WP-W 52  
DB 3 AILIPFLSLIPLTPQSAFAQSEBELKLESVVIVSRHGRVAPTKATQLMQDVTDPDAWPTW 62  
QY 53 PV---GLTPRGAALV-LLGY---RYFARGLL--PGCPAAG--TILADVDERTR-TGOAF 99  
DB 63 PVKLGWLTFRGELLAYLGHYQORLVADGLLAKKGCQSQGVAVIADVDERTRKTGEAF 122  
QY 100 AAGLAPGCAI-VH-----AGDDPIPHGLDTG-C-LDQA---DAILAAGEGGLT--ARHGL 146  
DB 123 AAGLAPDCAITVHTQADTSSPDPLFNPLKTVGCQDNLNANVTDAILSRAGGSIAFTGHRQ 182  
QY 147 T-----LAKVLNF--ASACLE-----PARVVLLAGHDTNLA-LGG-LDL-WQLP-OPDETTPGGALVFELWNR 271  
DB 183 TAFRELRVLPFPOSNCLKREKQDECSLTQALPSELKVSADNVSLTGAVSLASMLTEI 242  
QY 180 FLLEYAQPMVEVWGRI--SAEW-----LLHNAQF-LMNRTPYIA--RATPI--LIVTALS 228  
DB 243 FLQQAQGMPEPGWGRITDSHQWNTLLSLHNAQFYLLQRTPEVARSRAATPLLDLIKALT 302  
QY 229 -----PARVVLLAGHDTNLA-LGG-LDL-WQLP-OPDETTPGGALVFELWNR 271  
DB 303 PHPPQKQAYGVTLPSTVLFIAHDTNLANLGGALELWNTLPGQPDNTPPGGELVFERWR 362  
QY 272 -----YRVVMYQTMQDLRLNLEPLPRILP-----IPGC---GSEAACSLSDFARLV 314  
DB 363 LSDNSQWITQSVLFTQMRDKTFLSLNTPPGEVKLTLAGCEERNAQGMCSLAGTQIV 422  
QY 315 ----APAC 318  
DB 423 NEARIPAC 430

RESULT 2  
Q8GN88 PRELIMINARY; PRT; 432 AA.  
ID Q8GN88  
AC Q8GN88, 2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DE APPA.  
GN APPA.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX Zhang Z., He J., Yao B., Zhou Y., Chen Y., Yi Y.;  
RA "Production of phytase and acid phosphatase by use of silkworm-  
RT bioreactor";  
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX Chen Y., Zhu Z., Zhang Z., He J.;  
RA "Cloning and overexpression of phytase gene appA from Escherichia  
RT coli";  
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
DR GO; AF537219; AAN28334.1; -;  
DR GO; GO:0003993; F:acid phosphatase activity; IEA.  
DR InterPro; IPR000560; HisAc\_phsptase.  
DR Pfam; PF00328; acid phosphat; 1.  
DR PROSITE; PS00616; HIS ACID PHOSPHAT 1; 1.  
DR PROSITE; PS00778; HIS ACID PHOSPHAT 2; 1.  
SQ SEQUENCE 432 AA; 47055 MW; 55355D76E737737737 CRC64;

Query Match 37.0%; Score 609.5; DB 2; Length 432;  
Best Local Similarity 45.1%; Pred. No. 1.9e-38;  
Matches 193; Conservative 45; Mismatches 79; Indels 111; Gaps 36;

QY 2 SLLLAALALALAP-AAAAAP-LKLEKVVILSRHGRVSPTKA--LM--VSPD-WP-W 52  
DB 3 AILIPFLSLIPLTPQSAFAQSEBELKLESVVIVSRHGRVAPTKATQLMQDVTDPDAWPTW 62  
QY 53 PV---GLTPRGAALV-LLGY---RYFARGLL--PGCPAAG--TILADVDERTR-TGOAF 99  
DB 63 PVKLGWLTFRGELLAYLGHYQORLVADGLLAKKGCQSQGVAVIADVDERTRKTGEAF 122  
QY 100 AAGLAPGCAI-VH-----AGDDPIPHGLDTG-C-LDQA---DAILAAGEGGLT--ARHGL 146  
DB 123 AAGLAPDCAITVHTQADTSSPDPLFNPLKTVGCQDNLNANVTDAILSRAGGSIAFTGHRQ 182  
QY 147 T-----LAKVLNF--ASACLE-----PARVVLLAGHDTNLA-LGG-LDL-WQLP-OPDETTPGGALVFELWNR 271  
DB 183 TAFRELRVLPFPOSNCLKREKQDECSLTQALPSELKVSADNVSLTGAVSLASMLTKI 242  
QY 180 FLLEYAQPMVEVWGRI--SAEW-----LLHNAQF-LMNRTPYIA--RATPI--LIVTALS 228  
DB 243 FLQQAQGMPEPGWGRITDSHQWNTLLSLHNAQFYLLQRTPEVARSRAATPLLDLIKALT 302  
QY 229 -----PARVVLLAGHDTNLA-LGG-LDL-WQLP-OPDETTPGGALVFELWNR 271  
DB 303 PHPPQKQAYGVTLPSTVLFIAHDTNLANLGGALELWNTLPGQPDNTPPGGELVFERWR 362  
QY 272 -----YRVVMYQTMQDLRLNLEPLPRILP-----IPGC---GSEAACSLSDFARLV 314  
DB 363 LSDNSQWITQSVLFTQMRDKTFLSLNTPPGEVKLTLAGCEERNAQGMCSLAGTQIV 422  
QY 315 ----APAC 318  
DB 423 NEARIPAC 430

RESULT 3  
Q8RKD6 PRELIMINARY; PRT; 432 AA.  
ID Q8RKD6  
AC Q8RKD6, 2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DE Periplasmic phosphoanhydride phosphohydrolase.  
GN APPA.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90368616; PubMed=2168385;  
RA Dassa J., Macker C., Boquet P.-L.;  
RT "The complete nucleotide sequence of the Escherichia coli gene appA  
RT reveals significant homology between pH 2.5 acid phosphatase and  
RL glucose-1-phosphatase";  
RL J. Bacteriol. 172:5497-5500 (1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93054596; PubMed=1429631;  
RA Ostanin K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.M.,  
RA Van Etten R.L.;  
RT "Overexpression, site-directed mutagenesis, and mechanism of  
RT Escherichia coli acid phosphatase";  
RL J. Biol. Chem. 267:22830-22836 (1992).  
DR EMBL; I03374; AAA0006.1; -;  
DR GO; GO:0003993; F:acid phosphatase activity; IEA.  
DR InterPro; IPR000560; HisAc\_phsptase.  
DR Pfam; PF00328; acid phosphat; 1.  
DR PROSITE; PS00616; HIS ACID PHOSPHAT 1; 1.  
DR PROSITE; PS00778; HIS ACID PHOSPHAT 2; 1.  
KW Hydrolase.  
SQ SEQUENCE 432 AA; 46971 MW; AF86C41EA6193AC5 CRC64;

Query Match 36.9%; Score 607.5; DB 2; Length 432;  
 Best Local Similarity 45.1%; Pred. No. 2.7e-38;  
 Matches 193; Conservative 44; Mismatches 80; Indels 111; Gaps 36;

QY 2 SLLAALLAALAP-AAAAAP-LKLEKVVILSRHVSPTKA--LM--VSPD-WP-W 52  
 DB 3 AILIPFLSLLIPTQSAFAQSEPELKLESVVIVSRHGVAPTKATQMLQMDVTPDAWPTW 62

QY 53 PV---GLTPRGAALV-LIGY-----RYFARGLL--PGCPAAG--TILADVDETR-TGOAF 99  
 DB 63 PVKLGWLTFRGGLIAYLGHYQORLVADGLAKKGCPOGQVAVIADVDETRTKGEAF 122

QY 100 AAGLAPGCAI-VH-----AGDDPIPHGLDTG-C-LDOA---DAILAAGEGGLT--ARHGL 146  
 DB 123 AAGLAPDCAITVHTQADTSPPDLFNLKTVGVCOLDNANVTDAILSRAGSIADFTGHRQ 182

QY 147 T-----LAKYVNF--ASACLE-----CARVVGRLGPL-LASTLSEI 179  
 DB 183 TAFRELERVLPFQSNLCKREKQDESCSLTQALPSELKVSADNVSLTGAVSLASMLTEI 242

QY 180 FLLEYAOGPMVGVGRI--SAEW---LLHNAQF-LMNRTPVIA--RATPI--LIVTALS 228  
 DB 243 FLIQAQGMPEPGNGRIITDSHOWNTLLSLHNAQFYLLQRTPEVARSRATPLDLIKTALT 302

QY 229 -----PARVVLLAGHTNLA-LGG-IDL-WQLP-QDETTPPGALVFELWNR 271  
 DB 303 PHPPQKQAYGVTLPSTVLFIAGHDTNLANLGALSLNWTLPQDPNTPPGGELVFERRR 362

QY 272 -----YRVVMYQTMQDLNLEPLPRILP-----IPGC---GSEAACSLSDFARLV 314  
 DB 363 LSDNSQWQVSLVFTLQMDKTPLSLNTPPGEVKLTLAGCEERNAQCMCSLAGFTQIV 422

QY 315 ----APAC 318  
 DB 423 NEARIPAC 430

## RESULT 4

QSRK08 PRELIMINARY; PRT; 432 AA.  
 AC QSRK08;  
 DT 01-JUN-2002 (TREMELrel. 21, Created)  
 DT 01-JUN-2002 (TREMELrel. 21, Last sequence update)  
 DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)  
 DE Periplasmic phosphoanhydride phosphohydrolase.  
 GN APPA.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=90368616; PubMed=2168385;  
 RA Dassa J., Marck C., Boquet P.-L.L.;  
 RT "The complete nucleotide sequence of the Escherichia coli gene appA  
 RT reveals significant homology between pH 2.5 acid phosphatase and  
 RT glucose-1-phosphatase."  
 RL J. Bacteriol. 172:5497-5500(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=93054596; PubMed=1429631;  
 RA Ostanin K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.M.,  
 RA Van Etten R.L.;  
 RT "Overexpression, site-directed mutagenesis, and mechanism of  
 RT Escherichia coli acid phosphatase."  
 RL J. Biol. Chem. 267:22830-22836(1992).  
 DR EMBL; L03372; AAA00004.1; -  
 DR GO; GO:0003993; F:acid phosphatase activity; IEA.  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR InterPro; IPR000560; HisAc\_phsphtase.  
 DR Pfam; PF00328; acid\_phosphat\_1.  
 DR PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.  
 KW Hydrolase.

SQ SEQUENCE 432 AA; 46971 MW; 9A85536B57FCCFB5 CRC64;  
 Query Match 36.9%; Score 607.5; DB 2; Length 432;  
 Best Local Similarity 45.1%; Pred. No. 2.7e-38;  
 Matches 193; Conservative 44; Mismatches 80; Indels 111; Gaps 36;

QY 2 SLLAALLAALAP-AAAAAP-LKLEKVVILSRHVSPTKA--LM--VSPD-WP-W 52  
 DB 3 AILIPFLSLLIPTQSAFAQSEPELKLESVVIVSRHGVAPTKATQMLQMDVTPDAWPTW 62

QY 53 PV---GLTPRGAALV-LIGY-----RYFARGLL--PGCPAAG--TILADVDETR-TGOAF 99  
 DB 63 PVKLGWLTFRGGLIAYLGHYQORLVADGLAKKGCPOGQVAVIADVDETRTKGEAF 122

QY 100 AAGLAPGCAI-VH-----AGDDPIPHGLDTG-C-LDOA---DAILAAGEGGLT--ARHGL 146  
 DB 123 AAGLAPDCAITVHTQADTSPPDLFNLKTVGVCOLDNANVTDAILSRAGSIADFTGHRQ 182

QY 147 T-----LAKYVNF--ASACLE-----CARVVGRLGPL-LASTLSEI 179  
 DB 183 TAFRELERVLPFQSNLCKREKQDESCSLTQALPSELKVSADNVSLTGAVSLASMLTEI 242

QY 180 FLLEYAOGPMVGVGRI--SAEW---LLHNAQF-LMNRTPVIA--RATPI--LIVTALS 228  
 DB 243 FLIQAQGMPEPGNGRIITDSHOWNTLLSLHNAQFYLLQRTPEVARSRATPLDLIKTALT 302

QY 229 -----PARVVLLAGHTNLA-LGG-IDL-WQLP-QDETTPPGALVFELWNR 271  
 DB 303 PHPPQKQAYGVTLPSTVLFIAGHDTNLANLGALSLNWTLPQDPNTPPGGELVFERRR 362

QY 272 -----YRVVMYQTMQDLNLEPLPRILP-----IPGC---GSEAACSLSDFARLV 314  
 DB 363 LSDNSQWQVSLVFTLQMDKTPLSLNTPPGEVKLTLAGCEERNAQCMCSLAGFTQIV 422

QY 315 ----APAC 318  
 DB 423 NEARIPAC 430

## RESULT 5

QSRK08 PRELIMINARY; PRT; 432 AA.  
 AC QSRK08;  
 DT 01-JUN-2002 (TREMELrel. 21, Created)  
 DT 01-JUN-2002 (TREMELrel. 21, Last sequence update)  
 DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)  
 DE Periplasmic phosphoanhydride phosphohydrolase.  
 GN APPA.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=90368616; PubMed=2168385;  
 RA Dassa J., Marck C., Boquet P.-L.L.;  
 RT "The complete nucleotide sequence of the Escherichia coli gene appA  
 RT reveals significant homology between pH 2.5 acid phosphatase and  
 RT glucose-1-phosphatase."  
 RL J. Bacteriol. 172:5497-5500(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=93054596; PubMed=1429631;  
 RA Ostanin K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.M.,  
 RA Van Etten R.L.;  
 RT "Overexpression, site-directed mutagenesis, and mechanism of  
 RT Escherichia coli acid phosphatase."  
 RL J. Biol. Chem. 267:22830-22836(1992).  
 DR EMBL; L03370; AAA00002.1; -  
 DR GO; GO:0003993; F:acid phosphatase activity; IEA.  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR InterPro; IPR000560; HisAc\_phsphtase.  
 DR Pfam; PF00328; acid\_phosphat\_1.

DR PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.  
KW Hydrolase. 432 AA; 46971 MW; 5BBD632D4682EABF CRC64;  
SQ SEQUENCE

Query Match 36.9%; Score 607.5; DB 2; Length 432;  
Best Local Similarity 45.1%; Pred. No. 2.7e-38;  
Matches 193; Conservative 44; Mismatches 80; Indels 111; Gaps 36;

QY 2 SLILAALLAALAP-AAAAAP-LKLEKWTLSRHGVSPTKA--LM--VSPD-WP-W 52  
DB 3 AILIPFSLIPIPTPQSAFAQSEFELKESVIVSNGVRAPTKATQLMDVTPDAPTW 62  
QY 53 PV---GLTPRGAALV-LLGY---RYFARGLL--PGCPAAG--TILADVDETR-TGQAF 99  
DB 63 PVKLGWLTTPRGGLIAYLGHYQRLVADGLLAKGCPQSGQVAILADVDETRKTGEAF 122  
QY 100 AAGLAPGCAI-VH-----AGDDPIFHGLDTG-C-LDOA---DAILAAGEGGLT--ARHGL 146  
DB 123 AAGLAPGCAITVHTQADTSFDFLNFPLKTVGVQOLDNANVTDAISRAGGSADFTGHRQ 182  
QY 147 T-----LAKVINF--ASACLE-----CARVVGRLGFL--LASTLSEI 179  
DB 183 TAPRELSRVLPQSNCLKREKQDESCSLTQALPSELKVSADNVSLTGAVSLASMLTEI 242  
QY 180 FLLEYAQGPMEVGVGRI--SAEW---LLHNAQF-LMNRTPYIA--RATPI--LIVTALS 228  
DB 243 FLQQAQGMPEPGWGRITDSHQNTLLSHNAQFYLLQRTPEVARSRAPIPLDLIKTALT 302  
QY 229 -----PARVVLLAGHDTNLA-LGG-LDL-WQLP-OPDETPPGGALVPELWNR 271  
DB 303 PHPPQQAQYGVLTPTSFLVFIAGHDTNLANLGGALELNTWLPFGQPDNTPPGGELVFERWR 362  
QY 272 -----YRVVMYQMDQLNLEPLPRILP-----IPGC---GSEAACSLSDFARLV 314  
DB 363 LSDNSQWISVLFQTLQMRDKTFLSLNTPPGEVKLTLAGCEERNAQGMCSLAGTQIV 422  
QY 315 ----APAC 318  
DB 423 NEARIPAC 430

RESULT 6  
Q8KXD9 PRELIMINARY; PRT; 432 AA.  
ID Q8KXD9  
AC Q8KXD9  
DT 01-JUN-2002 (TremBLrel. 21, Created)  
DT 01-JUN-2002 (TremBLrel. 21, Last sequence update)  
DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)  
DE Periplasmic phosphoanhydride phosphohydrolase.  
GN APPA.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=9036816; PubMed=2168385;  
RA Dassa J., Marck C., Boquet P.-L.L.;  
RT "The complete nucleotide sequence of the Escherichia coli gene appa  
RT reveals significant homology between pH 2.5 acid phosphatase and  
RT glucose-1-phosphatase";  
RL J. Bacteriol. 172:5497-5500 (1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93054596; PubMed=1429631;  
RA Ostalan K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.M.,  
RA Van Etten R.L.;  
RT "Overexpression, site-directed mutagenesis, and mechanism of  
RT Escherichia coli acid phosphatase";  
RL J. Biol. Chem. 267:22830-22836 (1992).  
DR EMBL; L03371; AAA00003.1; -  
DR GO; GO:0003993; F:acid phosphatase activity; IEA.  
DR GO; GO:0016787; F:hydrolase activity; IEA.

DR InterPro; IPR000560; HisAc phsphtse.  
DR Pfam; PF00328; acid\_phosphat; 1.  
DR PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.  
KW Hydrolase. 432 AA; 47033 MW; 9F2B99DF9C368175 CRC64;  
SQ SEQUENCE

Query Match 36.8%; Score 606.5; DB 2; Length 432;  
Best Local Similarity 45.1%; Pred. No. 3.2e-38;  
Matches 193; Conservative 45; Mismatches 79; Indels 111; Gaps 36;

QY 2 SLILAALLAALAP-AAAAAP-LKLEKWTLSRHGVSPTKA--LM--VSPD-WP-W 52  
DB 3 AILIPFSLIPIPTPQSAFAQSEFELKESVIVSNGVRAPTKATQLMDVTPDAPTW 62  
QY 53 PV---GLTPRGAALV-LLGY---RYFARGLL--PGCPAAG--TILADVDETR-TGQAF 99  
DB 63 PVKLGWLTTPRGGLIAYLGHYQRLVADGLLAKGCPQSGQVAILADVDETRKTGEAF 122  
QY 100 AAGLAPGCAI-VH-----AGDDPIFHGLDTG-C-LDOA---DAILAAGEGGLT--ARHGL 146  
DB 123 AAGLAPGCAITVHTQADTSFDFLNFPLKTVGVQOLDNANVTDAISRAGGSADFTGHRQ 182  
QY 147 T-----LAKVINF--ASACLE-----CARVVGRLGFL--LASTLSEI 179  
DB 183 TAPRELSRVLPQSNCLKREKQDESCSLTQALPSELKVSADNVSLTGAVSLASMLTEI 242  
QY 180 FLLEYAQGPMEVGVGRI--SAEW---LLHNAQF-LMNRTPYIA--RATPI--LIVTALS 228  
DB 243 FLQQAQGMPEPGWGRITDSHQNTLLSHNAQFYLLQRTPEVARSRAPIPLDLIKTALT 302  
QY 229 -----PARVVLLAGHDTNLA-LGG-LDL-WQLP-OPDETPPGGALVPELWNR 271  
DB 303 PHPPQQAQYGVLTPTSFLVFIAGHDTNLANLGGALELNTWLPFGQPDNTPPGGELVFERWR 362  
QY 272 -----YRVVMYQMDQLNLEPLPRILP-----IPGC---GSEAACSLSDFARLV 314  
DB 363 LSDNSQWISVLFQTLQMRDKTFLSLNTPPGEVKLTLAGCEERNAQGMCSLAGTQIV 422  
QY 315 ----APAC 318  
DB 423 NEARIPAC 430

RESULT 7  
Q8KXD9 PRELIMINARY; PRT; 434 AA.  
ID Q8KXD9  
AC Q8KXD9  
DT 01-MAR-2002 (TremBLrel. 20, Created)  
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)  
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)  
DE Phosphoanhydride phosphorylase, pH 2.5 acid phosphatase,  
DE periplasmic.  
GN APPA OR Zi397 OR ECS1136.  
OS Escherichia coli O157:H7.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=83334;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
RX MEDLINE=21074935; PubMed=11206551;  
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
RA Rose D.J., Mayhew G.P., Evans P.S., Gregor J., Kirkpatrick H.A.,  
RA Postai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoukis K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blattner F.R.;  
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";  
RL Nature 409:529-533 (2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / RIMD 0509952;  
RX MEDLINE=21156231; PubMed=11258796;



RA	Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA	Ran C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tohe T.,
RA	Iida T., Takami H., Honda T., Sasekawa C., Ogasawara N., Yasunaga T.,
RA	Kuhara S., Shiba T., Hattori M., Shinagawa H.,
RT	"Complete genome sequence of enterohemorrhagic <i>Escherichia coli</i>
RT	O157:H7 and genetic comparison with a laboratory strain K-12.,"
RT	DNA Res. 8:11-22(2001).
DR	EMBL; AE005292; AAG55528.1; ALT_INIT.
DR	EMBL; AF002554; BAB34559.1; --.
DR	P1R; D85633; D85633.
DR	P1R; H90770; H90770.
DR	GO; GO:0003993; Fractid phosphatase activity; IEA.
DR	InterPro; IPR000560; HisAcid_phosphatse.
DR	Pfam; PF0328; acid_phosphat; 1.
DR	PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR	PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
DR	Complete proteome.
SK	SEQUENCE 434 AA; 47337 MW; F197DF7D1869F9C4 CRC64;

Query Match	36.8%;	Score	606.5;	DB	16;	Length	434;
Best Local Similarity	44.9%;	Pred. No.	3.2e-38;				
Matches	193;	Conservative	43;	Mismatches	81;	Indels	113;
Gaps	36;						
QY	2	SLLAALLAALAAPAAAAA----	EP-LKLEKVVILSRHGVRSPTKA--LM--VSPD-WP	51			
DB		: : : :	: : : :				
	3	AILPFLSLILPLTPQSAFQSQSBEPELKLSEVVIVSRHGVRATKATQLXQDVTTPDAMP	62				
QY	52	WPV-----GLTPRGAAVL-LIGY-----RYFARGLL--PGCPAAG--TILADVDETR--TQQ	97				
DB		: : : :	: : : :				
	63	NWPKLGLWTPRGGELIAYLGHYQORQLVADGLLTKKGCPOPGOVAIADVDETRTIGE	122				
QY	98	AFAAGLAPGCAI-VH-----AGDDPIPHGLDTG-C-LDOA---DAIIAAGGGTL--ARH	144				
DB		: : : :	: : : :				
	123	AFAAGLAPDCAITVHTQADTSSPPLEPNPLKTVGCQLDNANVTAILSRAGSIADTGTG	182				
QY	145	GLT----LAKVLNF--ASACLE-----CARVVRGLGLP--LASTLS	177				
DB		: : : :	: : : :				
	183	QTAFRELERVLNFPQSNCLNLRKQEDSCSLTQALPSFLKVSADNVSLTGAVSLASMLT	242				
QY	178	EIFLLEVAQGMVEYGCWRI--SASW----LHNAQF-LMNRTPVIA--RAPPI--LIVTA	226				
DB		: : : :	: : : :				
	243	EIFLLQQAQGMPEPGWGRITDShQWNTLSLHNAQFYLLQRTPEVASRATPLDLIMIA	302				
QY	227	LS-----PARVVLLAGHDNTILA-LGG-LDL-WOLP-QPDETTPGGALVPELW	269				
DB		: : : :	: : : :				
	303	LTPHPKQKQAGVTLPTSLVTFIAGHDNTILMGLGALINLTLPQPDNTPPGGLVFERW	362				
QY	270	NR-----YVRVMYCTMDLRNLEPLLPILP-----IPGC-----GSEACSLSDPAR	312				
DB		: : : :	: : : :				
	363	RLUSDNCWITQVSLVFTQLQMRDKTPTLSLNTPPGEVVKLTAGCEERNAQMCSLAGFTQ	422				
QY	313	LV-----APAC	318				
DB		: : :	: : :				
	423	IVNEARIPAC	432				
DB		: :	: :				

```

RESULT 8
Q8CW75      PRELIMINARY;      PRT;      445 AA.
ID          Q8CW75
AC          Q8CW75;
DT          01-MAR-2003 (T-EMBLrel. 23, Created)
DT          01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT          01-OCT-2003 (T-EMBLrel. 23, Last annotation update)
DE          Periplasmic appa protein precursor.
DE          APPA OR Cl121.
GN          GN
OS          Escherichia coli O6.
OC          Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC          Enterobacteriaceae; Escherichia.
OX          NCBI_TaxID=217992;
[1]
RN          SEQUENCE FROM N.A.
RP          STRAIN=O6:H1 / CFT073 / ATCC 700928;
RC          MEDLINE=22388234; PubMed=12471157;
RX          [1]

```

```

RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J.T., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
EL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
DR EMBL; AE016758; AA079589.1;
DR GO; GO:0003993; Fracid phosphatase activity; IEA.
DR InterPro; IPR000560; HisAc_phspatase.
DR Pfam; PF00328; acid_phosphat; 1.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT 1; 1.
DR DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
KW Complete proteome.
SQ SEQUENCE 446 AA; 48605 MW; F1308CD691DB9F49 CRC64;

Query Match 36.8%; Score 605.5; DB 16; Length 446;
Best Local Similarity 44.9%; Pred.No. 3.9e-38;
Matches 192; Conservative 44; Mismatches 81; Indels 111; Gaps 36;

QY 2 SLLLAALLAALAP-AAAAAP-LKLEKVVILSRHGVRSPTKA--LM--VSPD-WP-W 52
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 13 AILIPFLSLILPTPKSAFAQSPPELKLESVVIVSRHGVRAPTKATQLMQDVTPDAWPTW 72
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 53 PV---GLTPGGAALV-LLGY-----RVFARGLL---PGCPAAG--TILADVDERTR-TQAPF 99
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 73 FVKLGWLTPRGGELIAYLGHYQQRIVADGLLTKKGCPQGVVAISDVDERTRKTKGEAF 132
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 100 AAGLAPCAI-VH-----AGDDPIFGHLDTG-C-LQQA-----DAILAAGEGLT--ARHGL 146
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 133 AAGLAPDCAITVTDQTSSPDILFNLPTKGVQCLDNVNTDAILSPAGSIADFTCHRQ 192
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 147 T----LAKVNF--ASACLE-----CARVGRGLPL-LASTLSEI 179
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 193 TAPRELERVLPQSNLCNLRKQDSCSLTQALPSCLKVSADNVSLTGAVSLASMLTEI 252
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 180 FLLEYAQGPMEVCGWRI--SAEW---LIHNAOF-LMNRTPYIA--RATPI--LIVTALS 228
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 253 FLIQAQGMPEPGWGRITDSQHWNTLLSLHNAQFYLLQRTPEVARSRAPELLOLWAAIT 312
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 229 -----PARVVLGAHDTNLA-LGG-LDL-WQLP-OPDSTPPGALVFLWNR 271
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 313 PHSQKQAYGWTLPSTVLFITAGHDTNLANLGGALELNLTLPQGFNDTPPGSELVFERWR 372
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 272 -----YRVVMYQTMQDLNLEPLLRIPL-----IPGC---GSEACSLSDFARLV 314
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 373 LSDNSQIWQSLVFTLQIQQRKDTKPTSLNTPPGEVVKLTLAGCERNAQGCMSLAGTQIV 432
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 315 ----APAC 318
Db : : : : :
QY 433 NEARIPAC 440
Db : : : : :

```

```

RESULT 9
Q8RKD5
1D ID PRELIMINARY; PRT; 432 AA.
AC Q8RKD5;
DT 01-JUN-2002 (TRMBLrel. 21, Created)
DT 01-JUN-2002 (TRMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TRMBLrel. 24, Last annotation update)
DT 01-JUN-2003 (TRMBLrel. 24, Last annotation update)
DE Periplasmic phosphoanhydride phosphohydrolase.
GN APPA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_taxid=562;
RN [1]
RX SEQUENCE FROM N.A.
RY MEDLINE=90368616; PubMed=2168385;
RA Dasga J., Marck C., Boquet P.-L.I.;
RT "The complete nucleotide sequence of the Escherichia coli gene appA
RT reveals significant homology between pH 2.5 acid phosphatase and
RT glucose-1-phosphatase."

```

OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OC NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90368416; PubMed=2168385;  
RA Dassa J., Marck C., Boquet F.-L.L.;  
RT "The complete nucleotide sequence of the Escherichia coli gene ap  
RT reveals significant homology between pH 2.5 acid phosphatase and  
RT glucose-1-phosphatase.";  
RT

[illegible]

RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,  
RA Schwartz D.C., Blattner F.R.;  
RT "Complete genome sequence and comparative genomics of Shigella  
RT flexneri serotype 2a strain 2457T";  
RL Infect. Immun. 71:2775-2786(2003).  
DR EMBL; AF016981; AAP16495.1; -.  
SQ SEQUENCE 432 AA; 47136 MW; 200442239BC66DFC CRC64;  
  
Query Match 36.2%; Score 596.5; DB 16; Length 432;  
Best Local Similarity 44.6%; Pred. No. 1.8e-37;  
Matches 191; Conservative 43; Mismatches 83; Indels 111; Gaps 36;  
  
QY 2 SLLLAALLAALAP--AAAAAP-LKLVVILSRHGVSPFKA--LM--VSPD-WP-W 52  
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 62  
3 AILIFPLSLPLTQSAFAQSEPELKMESVIVSRHGVRAPTKATQLMQMVTDAWPTW  
QY 53 PV---GLTPRGAAALV-LLGY-----RYFARGLL--PGCP--AAGTILADVDRTR-TGOAF 99  
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 122  
63 PVKLGWLPFRGELLAYLGHVQRQLVADGLLAKKGCPSQAQVALIADVDRTRTKGEAF  
QY 100 AAGLAPGCAI-VH-----AGDDPIFHGLDTG-C-LDQA-----DAILAAGEGLT--ARHGL 146  
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 182  
123 TAGLAPDCAITVHTQADTSPPDLFNLKTVGCQLDNNANVTDAILCRAGGSIADFTGHRQ  
QY 147 T----LAKVLPF--ASACLE-----CARVVGRLGPI--LASTLSEI 179  
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 242  
183 TVFRELVNLPQSNCLNREKQDESCSLTQALPSELKVSADNVSLTGAVSLASMLTETI  
QY 180 FLLEYAQGPMEVWGRI--SAEW-----LLHNAQF-LMNRTPYIA--RATPI--LIVTALS 228  
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 302  
243 FLLOQAQGPMPGWRITDHSQWNTLLSLHNAQFYLLQRTPEVARSRAITPLDLIMALT  
QY 229 -----PARVVLLAGHDTNLA-LGG-LDL-WQLP-OPDETTPGGALVPELWNR 271  
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 362  
303 PHPPQKQAGVTLPTSVLFIAGHDNLNLAGLELNTLPGQPNTPPGGELVFERWR  
QY 272 -----YVRVMYQTMQRLNLEPLPRILP-----IPGC---GSEAAACSLSDPFLRV 314  
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 422  
363 LSDNSQWISVLFQTLQMRDKTFLSLNTPRGEVKLTLAGEERNAQGMCSLAGFTQIV  
QY 315 -----APAC 318  
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 430  
423 NEARIPAC 430  
  
RESULT 12  
Q8ZFP6 PRELIMINARY; PRT; 441 AA.  
AC Q8ZFP6  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Probable histidine acid phosphatase (EC 3.1.3.2) (Phosphoanhydride  
DE phosphorylase)  
GN YP01648 OR APPA OR Y1810.  
OS Yersinia pestis.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Yersinia.  
OX NCBI\_TaxID=632;  
RN [1] \_SEQUENCE FROM N.A.  
RP SEQUENCE FROM N.A.  
RC STRAIN=CO-32 / Biovar Orientalis;  
RX MEDLINE=21470413; PubMed=1586360;  
RA Farkhill J., Wren B.W., Thomson N.R., Tibball R.W., Holden M.T.G.,  
RA Prentice M.B., Sebailia M., James K.D., Churcher C., Mungall K.L.,  
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,  
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,  
RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,  
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,  
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;  
RT "Genome sequence of Yersinia pestis, the causative agent of plague";  
RL Nature 413:523-527(2001).  
RN [2]

RP SEQUENCE FROM N.A.  
RC STRAIN=KIMS / Biovar Mediaevalis;  
RX MEDLINE=22137863; PubMed=12142430;  
RA Dang W., Burland V., Plunkett G. III, Boutin A., Mayhew G.P., Liss P.,  
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,  
RA Petherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,  
RA Straley S.C., McDonough K.A., Milles M.L., Matson J.S., Blattner F.R.,  
RA Perry R.D.;  
RT "Genome sequence of Yersinia pestis KIM.";  
RL J. Bacteriol. 184:4601-4611(2002).  
DR EMBL; AJ414149; CAC90470.1; -.  
DR EMBL; AZ013783; AAM85378.1; -.  
DR PIR; AC0201; AC0201.  
DR GO; GO:0003993; F:acid phosphatase activity; IEA.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR000560; HisAc\_phsphtse.  
DR Pfam; PF00328; acid phosphat; 1.  
DR PROSITE; PS00616; HIS ACID PHOSPHAT\_1; 1.  
KW Hydrolase; Complete proteome.  
SQ SEQUENCE 441 AA; 47784 MW; CCD15COA64C98439 CRC64;  
  
Query Match 35.8%; Score 590; DB 16; Length 441;  
Best Local Similarity 43.4%; Pred. No. 5.9e-37;  
Matches 187; Conservative 42; Mismatches 88; Indels 114; Gaps 34;  
  
QY 1 VSLLAALLAAL--APAAAAAPLKLKVVILSRHGVSPFK--ALM--VSPD-WP-WP 53  
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 69  
10 LSLGLVLMGLAATAITAPVAEPGSGYTLERVVILSRHGVSPFKOTQMDVTFDKWPQW  
QY 54 VG---LTPRGAALVLL-----GYRYPARGLL-PGCPAAGTIL--ADVDRTR-TGOAF 101  
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 129  
70 VKAGYLTPRGAELVTLMGGFYGDYFRSLGLAAGCPAEGGVYAQADIDQRTLTGQAFLD  
QY 102 GLAPGCAI-VH-----AGDDPIFHGLDTG-C-LD--QADAILAAGEGG---LTARHGLT 147  
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 189  
130 GVAPGGGLTVHNOADLKKTDPLFHPVEAGVCKLDAQAOTKATIEEQGGGLDVTQSYAKP  
QY 148 LAK---VLNFAA--CLECAR-----VGRGLGPI--LASTLSEI 179  
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 249  
190 FAQMGDVLPFAAPSPCKSLQGGKTCDFAHFAANVNVNKGKTKVLSGFLASLTGEL  
QY 180 FLLEYAQGPMEVWGRI--SAEW-----LHNAQF-LMNRTPYIAR--ATPIL--LVTAL- 227  
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 309  
250 FLLOQAQGPMEVAVKQKGAENWVSLSLHNAQFNLMAKTPYIARHKGTPLQOQIDTALT  
QY 228 -----SPARVVLLAGHDTNLA-----LGGLDLWQLP-OPDETTPGGALVPELW 269  
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 368  
310 LQIDAQGGKLPISAQNRVFLGGHDNINIANIAGMLGAD-WQLPEQPDNTPGGGLVPELW  
QY 270 -----NRY--YVRVMYQTMQRLNLEPLPRILP-----IPGC---GSEAAACSLSDP- 310  
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 428  
369 QNPDNHQRYVAVKQFYQTMQRLNLEKLDLKNPAGIISVAVAGCENNGDDKLCELDTFQ  
QY 311 ---ARLVAPAC 318  
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 429  
429 KKVAVIEPAC 439  
  
RESULT 13  
Q9AAQ4 PRELIMINARY; PRT; 414 AA.  
AC Q9AAQ4  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Periplasmic phosphoanhydride phosphohydrolase.  
GN CC0542.  
OS Caulobacter crescentus.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;  
OC Caulobacteraceae; Caulobacter.  
OX NCBI\_TaxID=155892;  
RN [1] \_SEQUENCE FROM N.A.  
RP SEQUENCE FROM N.A.

```
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nieman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwin M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Kouri H., Shetty J., Berry K.,
RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.,
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL: AE005727; AA02529.1;
DR PIR: E87316; E87316.
DR HSP: P07102; IDKL.
DR TIGR: CC0542; -.
DR GO: GO:0003993; F:acid phosphatase activity; IEA.
DR GO: GO:0016787; F:hydrolase activity; IEA.
DR InterPro: IPR000560; HisAc phosphatase.
DR Pfam: PF00328; acid phosphatase.
DR PROSITE: PS00616; HIS ACID PHOSPHAT_1; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 414 AA; 44279 MW; 7105C64EA01C4051 CRC64;

Query Match 22.9%; Score 377.5; DB 16; Length 414;
Best Local Similarity 35.9%; Pred. No. 8.9e-21;
Matches 148; Conservative 47; Mismatches 112; Indels 105; Gaps 29;

QY 4 LLALLALALAPAAAEPLKLEKVVLSRHGVSPTKALMVSPD-----WP--- 51
DB 5 LRTAVVTALASAGAAAE--TLEKVVLSRHGVS-----AMSPERLEASARPFRFE 58
QY 52 WPGV-LTPRGAALVL-LG--YR--YFARGLL-FGCPAAGTILADVDEPT-RTQQAFAAGL 103
DB 59 VPAGHLTARGTTLVARVGDYRRHYAAQGLKPGDCASVYAVANVTQRTIATKAYRETL 118
QY 104 APGCAI-----VHAGD-PPHFGHLDTCGLDQADAILA-----AGE--GGLTARHGLTLAKVL 152
DB 119 APGCPVTNTVGEIGNIDEMFEPKAGIV-KADHALAPAAVAGRVGGDLTAWASHNQEA 177
QY 153 NFASACLECAR-----VVGRLGP-LLASTLSEIFILLEYAQG 187
DB 178 QLDALLMQCKGPCPPAPKRRVFDKPGFVGDGELAGLSPEAFASGVTESLLMAWADG 237
QY 188 P--MEVNGRISAE-----WLLHNAQF-LNRTPIYAR-----ATPILIVTALS 228
DB 238 RDFAGLWKSJDEEALTRSFTHQABFDLRLTPYVARTLAGHLADRLAATLRDGAAG 237
QY 229 P--ARVLLAGHDNTLA-LGGL--DLWOLP--QPDETTPGCGALVFELWN-----RYVRVM 276
DB 298 PVDARLVIIAGHDGTLASGLLXEWETLPGYQNPQPGALVFERNRDDGVRVVR 357
QY 277 Y--QTMQDLRLNLEPLRLP-----IPGCGSAA-----CELSDFARLVAPA 317
DB 358 FTGQSLSQLRNWNTALDAKTPPLSAFVFVQCGGTATPAFCRLEDFETVVRGA 409

RESULT 14
ID Q8P330 PRELIMINARY; PRT; 443 AA.
AC Q8P330;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Phosphoanhydride phosphohydrolase.
GN APPA OR XCC0760.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=2202145; PubMed=12024217;

da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furian L.R.,
da Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
da Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
da Cimarotte G., Cannavan F., Cardozo J., Chambergo F., Chapina L.P.,
da Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
da Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferri M.I.T.,
da Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
da Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.P.,
da Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
da Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
da Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
da Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
da Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
da Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
da Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
RL Nature 417:459-463(2002).
DR EMBL: AE012175; AA040075.1;
DR GO: GO:0003993; F:acid phosphatase activity; IEA.
DR GO: GO:0016787; F:hydrolase activity; IEA.
DR InterPro: IPR000560; HisAc phosphatase.
DR Pfam: PF00328; acid phosphatase.
DR PROSITE: PS00616; HIS ACID PHOSPHAT_1; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 443 AA; 46618 MW; 8C514434CC95867CE CRC64;

Query Match 21.8%; Score 359; DB 16; Length 443;
Best Local Similarity 33.8%; Pred. No. 2.5e-19;
Matches 135; Conservative 50; Mismatches 123; Indels 92; Gaps 27;

QY 1 VSLLLAALLAAL---APAAAABPLKLEKVVLSRHGVSPTKALMVS-----PD 49
DB 38 LGVLLVAALSGLCTPRPTFAASAAS-ATLERVVVVFRHGVRAPLQGEAAAHHYADAPWQ 96
QY 50 NWPVGL--TPRGAALVLGYY-----FARGLLP--GCPAAG--TILADVDEPT-RTQQA 98
DB 97 WSTPASLTVPGRKAVQLSGEYLQWLMOQLLRNGCPARGSLTWTANTQRTIDSCAL 156
QY 99 FAAGLAPGCAIVHAG-----DDPIFHGLDTCGLD-QADAILA-----GEGGLTARHG 145
DB 157 LADALAPGCN-VQAGHRAVGSNDPLFRPIEAGVFPDANAATAVAATQRTGTPAALQGHG 215
QY 146 L---TLAKVLPASACLECARVVGRL-----GPL-LASTLSEIFILLEYAQG-P 188
DB 216 AELQALQQLGCTKTPCFARMPSTLKFSADGRGLAGLPIDLASGTGTEVLLQYAEQWP 275
QY 189 M--EVNGRISAEWL-----LHNAQFLMNRTPY-----IARATPILIVTALS PARVV 233
DB 276 LSGVCGRATPERLAQVSRKHLALLFDIYAREPHMASRSCAPLAREVLQRFQDAGAP-KVS 334
QY 234 LLAGHDNTLA-----LGGLDLWOLP--QPDETTPGCGALVFELW-----NRYVRVMY--OTM 280
DB 335 VTVGSDTHIALUSLLGVH-FELPGYGADDPFGGALVGLWRTPDGORSVRAQYLAQSL 393
QY 281 DQLRLNLEPL-----LPRILPIPGCGSE--AACSLSDFA 312
DB 394 EQLRTVAPLDLDPHPLQQTALGVCAQGEQRMACPLPAFTQ 433

RESULT 15
ID Q8PFF53 PRELIMINARY; PRT; 435 AA.
AC Q8PFF53;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Acid phosphatase.
GN APPA OR XAC4132.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
```

[1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=306 / ATCC 13902 / XV 101;  
 RX MEDLINE=22022145; PubMed=12024217;  
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,  
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,  
 RA Canarotte G., Cannavan F., Cardoso J., Chambergo F., Ciapina L.P.,  
 RA Cicaralli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,  
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,  
 RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
 RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
 RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
 RA Martins E.C., Maidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,  
 RA Spinoia L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,  
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,  
 RA Setubal J.C., Kitajima J.P.;  
 RT "Comparison of the genomes of two Xanthomonas pathogens with differing  
 RT host specificities";  
 RL Nature 417:459-463(2002).  
 DR EXBL; AB012063; AM38967.1; -;  
 DR GO; GO:0003933; F:acid phosphatase activity; IEA.  
 DR InterPro; IPR000560; HisAc\_phosphatase.  
 DR Pfam; PF00328; acid\_phosphat; 1.  
 DR PROSITE; PS00616; HIS\_ACID\_PHOSPHAT\_1; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 435 AA; 46186 MW; F85EC7B63F1511B7 CRC64;

Query Match 21.4%; Score 353; DB 16; Length 435;  
 Best Local Similarity 34.3%; Pred. No. 7e-19;  
 Matches 147; Conservative 39; Mismatches 116; Indels 126; Gaps 31;

Qy	4	LLAALLAALAPAAA-AEPLKLVILSRHGVRSPTRKALMVSPD-----WPM- 52
Db	12	LLGLGLALVPLTPAQATDTPDLEKVVILKGVHVR-----AMSSPERLGRYSLHPHPF 67
Qy	53	--PVG--LTPGAALV--LIG--YR--YFARGLL--PGCPAAGTTLADVDERT-RTGQAPAA 101
Db	68	GVPAGHLTANGAQLERLFGDYRYAHYATLGLLNGDGNQA--YWANRTORTIASAQALAS 126
Qy	102	GLAPCAIVHAGDDPIFHGLDTGCLD-----QADAILAAGEGGLTARHG 145
Db	127	TLTFGCA-----DPVHH-VFTGSSDVLFDGTAALRQPQARMQAAIAGIGDAQWN 179
Qy	146	LTLAKVLN-FASACLECAR-----VVGRLL-----GPLL-----ASTL 176
Db	180	ATQSDALDTLEHLLQLCAQPCPAQAPGKRLTTPVAGLDDAGFSIPGLDGPAAAASGI 239
Qy	177	SEIFLEVAQGP--MEVCGHRISEAWLL-----HNAQFLMN-RTPYIAR--ATPI---LI 223
Db	240	TESLIMGWADGQDPFALGNQGLDEATLLRVFAPQAEFAULRAPTVARLASTPLAARLL 299
Qy	224	VTAL-----SPARVLLAGHDNLAL--GGLDL-WQLP--QPDETTPGGALV 265
Db	300	ATLQGGNDAAHAEADAIGDARLVVYSGHDGTLTLAGMFDLHMLPGVQSDQTVPGGALV 359
Qy	266	FELW-----NRYRVWY--QTMQDLRNLPLPRILP-----IPGCGSEAA---CSLSD 309
Db	360	FERRWRADGGERVIRLRYTAQSLTLRERVPLTLOAPPSPSAIFPGCSTATPAYDCLPQ 419
Qy	310	PARLVAPA 317
Db	420	PARIVQAA 427

Search completed: April 30, 2004, 12:41:28  
 Job time : 31.8084 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2004, 19:58:17 ; Search time 3538.44 Seconds  
(without alignments)  
16230.166 Million cell updates/sec

Title: US-10-021-723B-3

Perfect score: 1325

Sequence: 1 atgtcagtattagaaatcg.....aacctgctgccacatctaa 1325

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*

1: gb.ba:\*

2: gb.htg:\*

3: gb.in:\*

4: gb.ov:\*

5: gb.pat:\*

6: gb.ph:\*

7: gb.pl:\*

8: gb.pr:\*

9: gb.ro:\*

10: gb.ats:\*

11: gb.un:\*

12: gb.vi:\*

13: gb.ba:\*

14: gb.fun:\*

15: gb.in:\*

16: gb.mu:\*

17: gb.ov:\*

18: gb.pat:\*

19: gb.ph:\*

20: gb.pl:\*

21: gb.pr:\*

22: gb.ro:\*

23: gb.ats:\*

24: gb.un:\*

25: gb.vi:\*

26: gb.ba:\*

27: gb.fun:\*

28: gb.in:\*

29: gb.mu:\*

30: gb.ov:\*

31: gb.pat:\*

32: gb.ph:\*

33: gb.pl:\*

34: gb.pr:\*

35: gb.ro:\*

36: gb.ats:\*

37: gb.un:\*

38: gb.vi:\*

39: gb.ba:\*

40: gb.fun:\*

41: gb.in:\*

42: gb.mu:\*

43: gb.ov:\*

44: gb.pat:\*

45: gb.ph:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1325	100.0	1325	6	AX556791	AX556791 Sequence
2	1306	98.6	10029	1	AE013783	AE013783 Yersinia
3	1306	98.6	193050	1	AJ414149	AJ414149 Yersinia
4	1292.2	97.5	1325	6	AX556789	AX556789 Sequence
5	827.6	62.5	1326	6	AX556799	AX556799 Sequence
6	323	24.4	2558	1	AY378096	AY378096 Obesumbac
7	321.4	24.3	2161	1	AY425982	AY425982 Obesumbac
8	225.2	17.1	5421	6	AX042377	AX042377 Sequence
9	225.2	17.0	1901	1	EC0ACPHD	EC0ACPHD Escherichia
10	225.2	17.0	1901	1	EC0ACPHD	EC0ACPHD Escherichia
11	225.2	17.0	1901	1	EC0ACPHD	EC0ACPHD Escherichia
12	225.2	17.0	1901	6	AX356571	AX356571 Sequence
13	225.2	17.0	1901	6	AX356572	AX356572 Sequence
14	224.6	17.0	3470	6	AX042376	AX042376 Sequence
15	224.6	17.0	4060	6	AX042374	AX042374 Sequence
16	224.6	17.0	6116	6	AX042375	AX042375 Sequence
17	224.6	17.0	6708	6	AX042373	AX042373 Sequence
18	224.6	17.0	10784	1	AE000200	AE000200 Escherich
19	224.6	17.0	17732	6	AX042378	AX042378 Sequence
20	224.6	17.0	18882	1	D90735	D90735 Escherichia
21	224.6	17.0	20623	6	AX042372	AX042372 Sequence
22	223.6	16.9	1901	1	EC0ACPHA	EC0ACPHA Escherichia
23	223.6	16.9	1901	1	EC0ACPHA	EC0ACPHA Escherichia
24	223.6	16.9	1901	1	EC0ACPHA	EC0ACPHA Escherichia
25	223.6	16.9	1901	1	EC0ACPHA	EC0ACPHA Escherichia
26	222	16.8	1901	1	EC0ACPHC	EC0ACPHC Escherichia
27	222	16.8	1901	1	EC0ACPHC	EC0ACPHC Escherichia
28	222	16.8	1901	1	EC0ACPHF	EC0ACPHF Escherichia
29	218.8	16.5	1323	6	AR108133	AR108133 Sequence
30	218.8	16.5	1323	6	AR127818	AR127818 Sequence
31	218.8	16.5	1323	6	AR130956	AR130956 Sequence
32	218.8	16.5	1323	6	AR130956	AR130956 Sequence
33	218.8	16.5	1323	6	AR130956	AR130956 Sequence
34	218.8	16.5	1323	6	AR130956	AR130956 Sequence
35	218.2	16.5	1489	6	AR277753	AR277753 Sequence
36	218.2	16.5	10029	1	AE005292	AE005292 Escherich
37	218.2	16.5	32773	1	AP002554	AP002554 Escherich
38	215	16.2	11761	1	AE015127	AE015127 Shigella
39	215	16.2	292504	1	AE016981	AE016981 Shigella
40	211.8	16.0	301276	1	AE016758	AE016758 Escherich
41	177.4	13.4	1431	6	AX556797	AX556797 Sequence
42	143.6	10.8	1272	6	BD074747	BD074747 Novel phy
43	68.4	5.2	1675	1	EC0AGPA	EC0AGPA Escherich
44	68.4	5.2	11458	1	AE015129	AE015129 Shigella
45	68.4	5.2	11710	1	AE000202	AE000202 Escherich

# ALIGNMENTS

RESULT 1	AX556791	AX556791	1325 bp	DNA	linear	PAT 27-NOV-2002
LOCUS	Sequence 3 from Patent WO0248332.					
DEFINITION	Sequence 3 from Patent WO0248332.					
ACCESSION	AX556791					
VERSION	AX556791.1	GI:258999903				
KEYWORDS	Yersinia pestis					
SOURCE	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Yersinia.					
ORGANISM	Yersinia pestis					
REFERENCE	1					
AUTHORS	Short, J., Mathur, E.J., Richardson, T., Robertson, D. and Barton, N.					
TITLE	Recombinant phytases and uses thereof					
JOURNAL	Patent: WO 0248332-A 3 20-JUN-2002;					

FEATURES  
source  
DIVERSA CORPORATION (US)  
Location/Qualifiers  
1. 1325  
/organism="Yersinia pestis"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:632"

ORIGIN  
Query Match 100.0%; Score 1325; DB 6; Length 1325;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGTCAGTATTAGAAATCGTACGGCTATCTGGCTGTATGATGCTAAAGCGGATTG 60  
Db 1 ATGTCAGTATTAGAAATCGTACGGCTATCTGGCTGTATGATGCTAAAGCGGATTG 60  
Qy 61 GCTGCTATTACTGGCGGTAGCGCGGAGCATCGGCTATCTTTAGAAAGTGTGTT 120  
Db 61 GCTGCTATTACTGGCGGTAGCGCGGAGCATCGGCTATCTTTAGAAAGTGTGTT 120  
Qy 121 ATTTGAGTCCCATGGTGTCTCGCCGACCAACACGACGCTTATGAATGATGTTA 180  
Db 121 ATTTGAGTCCCATGGTGTCTCGCCGACCAACACGACGCTTATGAATGATGTTA 180  
Qy 181 CGCAGATAGTGGCCACATGCGGTAAAGCGGGTATTAAACGCACTGTGTGGG 240  
Db 181 CGCAGATAGTGGCCACATGCGGTAAAGCGGGTATTAAACGCACTGTGTGGG 240  
Qy 241 AGTTGGTCAATTGATGGGGGGTTTATGGTGATTAATTCGACGCTTGTGTTAG 300  
Db 241 AGTTGGTCAATTGATGGGGGGTTTATGGTGATTAATTCGACGCTTGTGTTAG 300  
Qy 301 CGCGGGATGTCGGCAGAGGGGGGTATATGCAAGCAGATATCGATCAACGTACCC 360  
Db 301 CGCGGGATGTCGGCAGAGGGGGGTATATGCAAGCAGATATCGATCAACGTACCC 360  
Qy 361 GCTTAACGACAGGATTTCTTGATGCTGCTCGCGGTGCTGCTGATGCGTGCATA 420  
Db 361 GCTTAACGACAGGATTTCTTGATGCTGCTCGCGGTGCTGCTGATGCGTGCATA 420  
Qy 421 ATCAGCGGATTTGAAAGACCGATCCCTGTTCATCGGTAGAGATGCGTGTGTA 480  
Db 421 ATCAGCGGATTTGAAAGACCGATCCCTGTTCATCGGTAGAGATGCGTGTGTA 480  
Qy 481 AGTTAGACACGACACAGATAGAGATGAGATGAGATGAGATGAGATGAGATGAG 540  
Db 481 AGTTAGACACGACACAGATAGAGATGAGATGAGATGAGATGAGATGAGATGAG 540  
Qy 541 CGGTGAGCCAGCGCTAGCTAAACCTTTTCCAGATGGGGACGCTGCAATTTTGGG 600  
Db 541 CGGTGAGCCAGCGCTAGCTAAACCTTTTCCAGATGGGGACGCTGCAATTTTGGG 600  
Qy 601 CTTCTCTTATTCGAAATCTTTGACAGCAGGAAAGCTGTGATTTGGCCACTTTG 660  
Db 601 CTTCTCTTATTCGAAATCTTTGACAGCAGGAAAGCTGTGATTTGGCCACTTTG 660  
Qy 661 CGGCCAATGAAGTTAAGCTTAATAAGAGGAGGACAAAGTGACCTCAGTGGGCCACTGG 720  
Db 661 CGGCCAATGAAGTTAAGCTTAATAAGAGGAGGACAAAGTGACCTCAGTGGGCCACTGG 720  
Qy 721 CATTATCATGCATTTGGGTGAATCTTTTATTAACAGCAGCAGGATGCGAGAG 780  
Db 721 CATTATCATGCATTTGGGTGAATCTTTTATTAACAGCAGCAGGATGCGAGAG 780  
Qy 781 TTGCTGGCAACCGCTAAAGGGGCGAGATTTGGGTATCCTTTATTTGATACATAAGC 840  
Db 781 TTGCTGGCAACCGCTAAAGGGGCGAGATTTGGGTATCCTTTATTTGATACATAAGC 840  
Qy 841 CGCAATTAATTAATGGCAAAACACCGGTATATCGCCGTATATAAGGAGCGCCATTAT 900  
Db 841 CGCAATTAATTAATGGCAAAACACCGGTATATCGCCGTATATAAGGAGCGCCATTAT 900  
Qy 901 TACAGCAGATAGATCGGCTTTTAACTTCACTGGATGCTCAGGGGCAAGCTACCCA 960

Db 901 TACAGCAGATAGATCGGCTTTTAACTTCACTGGATGCTCAGGGGCAAGCTACCCA 960  
Qy 961 TTTTCCAGCCCAAAACCGGGTCTTTTCTCGGTGGGCAATGATACCAATATTTGCCAATATTG 1020  
Db 961 TTTTCCAGCCCAAAACCGGGTCTTTTCTCGGTGGGCAATGATACCAATATTTGCCAATATTG 1020  
Qy 1021 CGGATGCTGGGAGCGGCTGCGAGCTACCGAGCACTGATTAATCTCCACCGGTTG 1080  
Db 1021 CGGATGCTGGGAGCGGCTGCGAGCTACCGAGCACTGATTAATCTCCACCGGTTG 1080  
Qy 1081 GGGGATTTGGTTTTTGAATATGCAAAATCGGATAACCCAGCGCTAGCTTGGCGTGA 1140  
Db 1081 GGGGATTTGGTTTTTGAATATGCAAAATCGGATAACCCAGCGCTAGCTTGGCGTGA 1140  
Qy 1141 AGATGTTCTACCAACGATGATGCTGGTATGCGGAAATTTGGATCTGAAAAATA 1200  
Db 1141 AGATGTTCTACCAACGATGATGCTGGTATGCGGAAATTTGGATCTGAAAAATA 1200  
Qy 1201 ACCCAGCGGGTATTATTTCGGTTGCACTGCTGTTGTAATAAATACCGGTGACGATAAGC 1260  
Db 1201 ACCCAGCGGGTATTATTTCGGTTGCACTGCTGTTGTAATAAATACCGGTGACGATAAGC 1260  
Qy 1261 TTTTCCGAGCTTGATACATTTCCAAAGAAAGTGGCTAAGGTAATGAACCTGCTGCCACA 1320  
Db 1261 TTTTCCGAGCTTGATACATTTCCAAAGAAAGTGGCTAAGGTAATGAACCTGCTGCCACA 1320  
Qy 1321 TCTAA 1325  
Db 1321 TCTAA 1325

RESULT 2  
AE013783  
LOCUS  
Yersinia pestis KIM section 183 of 415 of the complete genome.  
DEFINITION  
AE013783 AE009952  
ACCESSION  
AE013783.1 GI:21958612  
VERSION  
KEYWORDS  
SOURCE  
Yersinia pestis KIM  
Yersinia pestis KIM  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Yersinia.  
REFERENCE  
1 (bases 1 to 10029)  
Deng, W., Burland, V., Plunkett, G. III, Boutin, A., Mayhew, G. F.,  
Liss, P., Perna, N. T., Rose, D. J., Mau, B., Zhou, S., Schwartz, D. C.,  
Fetherston, J. D., Lindler, L. E., Brubaker, R. R., Plana, G. V.,  
Straley, S. C., McDonough, K. A., Nilles, M. L., Matson, J. S.,  
Blattner, F. R. and Perry, R. D.  
Genome Sequence of Yersinia pestis KIM  
J. Bacteriol. 184 (16), 4601-4611 (2002)  
12142430  
2 (bases 1 to 10029)  
Deng, W., Burland, V., Plunkett, G. III, Boutin, A., Mayhew, G. F.,  
Liss, P., Perna, N. T., Rose, D. J., Mau, B., Zhou, S., Schwartz, D. C.,  
Fetherston, J. D., Lindler, L. E., Brubaker, R. R., Plana, G. V.,  
Straley, S. C., McDonough, K. A., Nilles, M. L., Matson, J. S.,  
Blattner, F. R. and Perry, R. D.  
Direct Submission  
Submitted (21-FEB-2002) Genetics, University of Wisconsin, 445  
Henry Mall, Madison, WI 53706, USA  
Location/Qualifiers  
1. 10029  
/organism="Yersinia pestis KIM"  
/mol\_type="genomic DNA"  
/strain="KIM"  
/db\_xref="taxon:187410"  
complement(104..1147)  
/gene="y1803"  
complement(104..1147)  
/gene="y1803"  
/function="regulator; degradation of small molecules;  
Carbon compounds"

FEATURES  
source  
genetic  
CDS



/note="residues 1 to 330 of 347 are 60.00 pct identical to residues 1 to 330 of 331 from GenPept : >gb|AA057489.1|AE005467.8 (AE005467) sucrose specific transcriptional regulator [Escherichia coli O157:H7 EDL933]"  
/codon\_start=1  
/transl\_table=1  
/product="putative sucrose specific transcriptional regulator"  
/protein\_id="AA05371.1"  
/db\_xref="GI:21958613"  
/translation="MVSLEKVNALSVLMTASRAINSPEQLRPETVQYVMAIESLN YVPDFAARKRGHATKSTLAVLAWDTATPFSVILLSTELTAREFGWNSFLVNLTS REDSARLNOLLAPQDGIIFTAGLRQVTIPEFRDKNIVLANCVSESHSIIPSD DFGQYQAKVLLQGGYQRPCLVLPCTCKAGKARAGVEKAWQESALPTEQLNOCHI ANGDEHGVQIVPLLERCPKGNFNVLCIGDDRIAFVLAQVLLAGKVIPIQGVAVIG XGDHVMGSELYLPPLTTVQLPYDIGHQQAALLIRGESAIRHLPCPLDERGSIERS MPNNCVPLFSKQR"  
/complement(1423. .1683)  
/gene="y1804"  
/complement(1423. .1683)  
/gene="y1804"  
/function="unknown"  
/note="residues 37 to 82 of 86 are 32.60 pct identical to residues 617 to 659 of 1018 from GenPept : >gb|AAK31959.1|AF285580.1 (AF285580) serine/threonine kinase 31 [Mus musculus]"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical"  
/protein\_id="AA05372.1"  
/db\_xref="GI:21958614"  
/translation="MSSINQTSVRAHTGETGALNPGEDHIIAHEALLVALMVECANES DNDQFLSIRETLNGILSSRYLDANAIAINGLIDEADKANR"  
/complement(1803. .3440)  
/gene="aer"  
/note="synonym: y1805"  
/complement(1803. .3440)  
/gene="aer"  
/function="regulator; degradation of small molecules; Carbon compounds"  
/note="residues 26 to 533 of 545 are 47.44 pct identical to residues 7 to 504 of 506 from E. coli K12 : B3072; residues 20 to 535 of 545 are 53.68 pct identical to residues 1 to 514 of 514 from GenPept : >emb|CAD18375.1| (AL646083) probable aerotaxis sensor receptor (chemotaxis transducer) transmembrane protein [Ralstonia solanacearum]"  
/codon\_start=1  
/transl\_table=1  
/product="aerotaxis sensor receptor, flavoprotein"  
/protein\_id="AA05373.1"  
/db\_xref="GI:21958615"  
/translation="MKYNGSIDYQGSPPFTTEVMNRVKNKPYRQYPIERDITLQSTTD IHGNIAYANAAPVRAAGFPEYQELQGGHVVHRPMPAPAFADMMQTNLAGLWSSIV KNRKNGDYVVRANATPLRNGRLTGIVSIAPTRDEVKQAEALYDFNTGAKKR HIALYRGILVETGWSLPLSFQTLPLRWLSALLSSAIIPTAASVMEVAGOPLLIIL GSVALTWSVLASLWLERQIARPIAIIQQAQDVSSGAGDYVQLNRVDEIGYLMRSVN QLGNLRSLTDVGVQDGINFASNEIAAGNRELKVRTEQATDNLQHVIVSATEQLAAT VONSANANETTSIAEMSHAAKSGELMHQVETMGITINDSHRVIDIISVIEGIAP QTNILALNAEVAEQAQGFQFVAVASEVRLAQRSSTRAKEIKHIIETSVERRDG STLVQSGAATMDNIVKQASQVSTLISEISTSTHTQALQIQRSIRLQDQTYQNA MVEQYAGAAEALHRLTLTAAYRIYRLNNVGLRD"  
/complement(4303. .5160)  
/gene="y1806"  
/complement(4303. .5160)  
/gene="y1806"  
/function="putative"  
/note="residues 1 to 279 of 285 are 65.59 pct identical to residues 1 to 279 of 286 from E. coli K12 : B4211; residues 1 to 282 of 285 are 65.60 pct identical to residues 1 to 282 of 284 from GenPept : >emb|CAB69475.1| (A92090) ORF1 [Pseudomonas sp.]"

/codon\_start=1  
/transl\_table=1  
/product="putative oxidoreductase"  
/protein\_id="AA05374.1"  
/db\_xref="GI:21958616"  
/translation="MIAVTGATGQLRVINVLKKVPAHDIIAAVRSPDKAKLMAL GYVQADYNQPTTLDISAFKGVGKLLIISSEVQORVAQHTAVNAAKRAGVGLIAYT SULHADKSPGLGHEHRAETGLLRDSGLPVVLLRNGWYTYENYASIAIPALAHGAFIGA AGRGFIASAREDYAEAAATGLLVQENQAGKIYVLAGDDAYTLSEFSAITRQSGKAVA YQDLSEADFKHALISAGLPDGLAGLADSDAGAAQGGLPDDSHLTKLIGRPTTTPYAK VITATLATR"  
/complement(5131. .5703)  
/gene="y1807"  
/complement(5131. .5703)  
/gene="y1807"  
/note="residues 68 to 189 of 190 are 62.29 pct identical to residues 34 to 155 of 156 from E. coli K12 : B4212"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="AA05375.1"  
/db\_xref="GI:21958617"  
/translation="MAGTGCDDNDADPVEEQLTLREDYGVICPKSSLKPSVAYI SLIVSTYQKVSVRVTKLQNMGNISFPQLRGELLNVDCPSREILKVTWRSWGLVL IALSNETURFSAIRKIGGVSEKMLAQIQLHLEDGFDIDRIAYFVVPFVHYSLTPLG KBIQOVVEGLTVWLETNLHRIMEQRQRSV"  
/complement(5785. .5934)  
/gene="y1808"  
/complement(5785. .5934)  
/gene="y1808"  
/function="unknown"  
/note="residues 6 to 49 of 49 are 42.22 pct identical to residues 61 to 101 of 380 from GenPept : >dbj|BAE70104.1| (AP002924) cytochrome b [Lampris guttatus]"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical"  
/protein\_id="AA05376.1"  
/db\_xref="GI:21958618"  
/translation="MLEGTAPDVLPTIGQARFGLRHWLKSTAPTAGFPFICIYL SAARG"  
/complement(5931. .6023)  
/gene="y1809"  
/complement(5931. .6023)  
/gene="y1809"  
/function="unknown"  
/note="residues 1 to 26 of 30 are 34.61 pct identical to residues 16 to 41 of 444 from GenPept : >gb|AAK78422.1|AE007558.12 (AE007558) Permease, putative chloride channel [Clostridium acetobutylicum]"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical"  
/protein\_id="AA05377.1"  
/db\_xref="GI:21958619"  
/translation="MSFTWVILGVVLGCVIGALLKIFGKKKDR"  
/complement(6235. .7560)  
/gene="appa"  
/note="synonym: y1810"  
/gene="appa"  
/function="enzyme; central intermediary metabolism: Pool, multipurpose conversions"  
/note="pH 2.5 acid phosphatase; periplasmic; residues 13 to 441 of 441 are 45.87 pct identical to residues 5 to 432 of 432 from E. coli K12 : B0980"  
/codon\_start=1  
/transl\_table=1  
/product="phosphoanhydride phosphorylase"  
/protein\_id="AA05378.1"  
/db\_xref="GI:21958620"  
/translation="MSVLNRVRLSGLVLMLSGLAAITAPVAAPSGYTLERVILSR HGVRSTKQTQLMNDVTPDKWQWPVKAGVLTARGAELVTLMGGFYDGRSLGLIAA

GCPAEGVYQAQIDQTRITGQAFIDGVAPOGLTVHNOADLKEKTDLPHFVEAGVC  
 KLDAAQTDRAIEQLGPDLDTSQRYAKPFAQMGVDVNFPAASPYCKSLQOQKTCDFP  
 HFAAEVNVNKEGTQVTLSPALSLTGEIPLFLOAQAQPEVAVCRKLKRAENWYSL  
 SLHNAQFNLMKATPYTARHKGTPFLQOIIDTALTLODAQOGLP.SAOKRVLFLGSHD  
 TNFANAGMLGADWOLPEOPDNTPPGGGLVFEELMQNPONHORYVAVKWPQYMDQLRN  
 AEKLDLKNPAGIISVAVAGCENNDKLCBLDTFOKKVAKVIEPACHI'  
 complement (7862..8077)  
 /gene="y1811"  
 /complement (7862..8077)  
 /notes="residues 9 to 71 of 71 are 41.26 pct identical to  
 residues 4 to 66 of 67 from E. coli K12 : B1541; residues  
 6 to 70 of 71 are 44.61 pct identical to residues 1 to 65  
 of 67 from Genpept : >emb|CAD09963.1| (AUS1384)  
 hypothetical protein [Salmonella enterica subsp. enterica]

Query Match 98.6%; Score 1306; DB 1; Length 10029;  
 Best Local Similarity 99.5%; Pred. No. 0;  
 Matches 1320; Conservative 0; Mismatches 5; Indels 1; Gaps 1;  
 1 ATGTCAGTATTAGAAAATCGTGACGGCTATCTGGCTGGTATTGATGTAAGCGGATTG 60  
 6235 ATGTCAGTATTAGAAAATCGTGACGGCTATCTGGCTGGTATTGATGTAAGCGGATTG 6294  
 61 GCTGCTATTACTGCGCGGTAGCGCGGAGCCATCGGGCTATCTTTAGAAAGTGTGGTT 120  
 6295 GCTGCTATTACTGCGCGGTAGCGCGGAGCCATCGGGCTATCTTTAGAAAGTGTGGTT 6354  
 121 ATTTGAGTGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 179  
 6355 ATTTGAGTGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6414  
 180 ACCCCAGATAAGTGGCCACAAATCGCGGTAAAGCGGGTATTATACGCCACGCTGTGGC 239  
 6415 ACCCCAGATAAGTGGCCACAAATCGCGGTAAAGCGGGTATTATACGCCACGCTGTGGC 6474  
 240 GAGTGTGTCATGATGAGGGGGTTTTATGTTGATTAATTTGCGAGCTTGGTTGTTA 299  
 6475 GAGTGTGTCATGATGAGGGGGTTTTATGTTGATTAATTTGCGAGCTTGGTTGTTA 6534  
 300 GCGCGCGGATGTCGCGCAGAGGGGGGTATATGACAGCAGATATCGATCAACGTACC 359  
 6535 GCGCGCGGATGTCGCGCAGAGGGGGGTATATGACAGCAGATATCGATCAACGTACC 6594  
 360 CGTTTAAACCGCAGCAGCATTTCTTGATGTTGGCTCGCGGGTGTGGTTGACCGTGAT 419  
 6595 CGTTTAAACCGCAGCAGCATTTCTTGATGTTGGCTCGCGGGTGTGGTTGACCGTGAT 6654  
 420 AATCAGCGCATTTGAAAGACCGATCCCTGTTCCATCCGTAGACGCTGCGGTGT 479  
 6655 AATCAGCGCATTTGAAAGACCGATCCCTGTTCCATCCGTAGACGCTGCGGTGT 6714  
 480 AAGTTAGACAAACACACAAACAGATAAAGCGATTGAAGACGATTGGCGGGCGGTTAGAT 539  
 6715 AAGTTAGACCGCCACACAAACAGATAAAGCGATTGAAGACGATTGGCGGGCGGTTAGAT 6774  
 540 ACGGTGAGCGCAGCTACGTAAACCTTTGCCCAGATGGGACGCTGCTGAATTTGGC 599  
 6775 ACGGTGAGCGCAGCGCTACGTAAACCTTTGCCCAGATGGGACGCTGCTGAATTTGGC 6834  
 600 GCTTCTCTTATTCGAATCTTTGCAACAGCAGAGGAAACGTTGATTTGCCACTTT 659  
 6835 GCTTCTCTTATTCGAATCTTTGCAACAGCAGAGGAAACGTTGATTTGCCACTTT 6894  
 660 GCGGCCAATGAAGTTAAGTTAATAAGAGGACAAAGTGAACCTTCAGTGGGCCACTG 719  
 6895 GCGGCCAATGAAGTTAAGTTAATAAGAGGACAAAGTGAACCTTCAGTGGGCCACTG 6954  
 720 GCGTTATCATCGCATTTGGGTGAATCTTCTTATTACAAAACGCAAGCCATGCCAGAG 779  
 6955 GCGTTATCATCGCATTTGGGTGAATCTTCTTATTACAAAACGCAAGCCATGCCAGAG 7014  
 780 GTTGCCCTGGCAACCGCTTAAAGGGGGGAGAAATGGGTATCTCTTATGTCATTACATAAC 839

7015 GTTGCCCTGGCAACCGCTTAAAGGGGGGAGAAATGGGTATCTCTTATTTGTCATTACATAAC 7074  
 840 GCGCAATTTAATTTAATGCGCAAAACACCGTATATCCCGTGCATAAAGGAGCGCCATT 899  
 7075 GCGCAATTTAATTTAATGCGCAAAACACCGTATATCCCGTGCATAAAGGAGCGCCATT 7134  
 900 TTACAGCAGATAGATACCGCTTTAAACCCCTTCAACTGGATGCTCAGGGGGCAAAAGCTACCC 959  
 7135 TTACAGCAGATAGATACCGCTTTAAACCCCTTCAACTGGATGCTCAGGGGGCAAAAGCTACCC 7194  
 960 ATTTTCAGCCCAAAACCGGGTCTTCTTCTCGGTGGGATGATACCAATATTTGCCAATATT 1019  
 7195 ATTTTCAGCCCAAAACCGGGTCTTCTTCTCGGTGGGATGATACCAATATTTGCCAATATT 7254  
 1020 GCGGGTATGCTGGAGCGGACTGCGAGTACCCGAGCAACCTGATATATATCTCCACCAAGT 1079  
 7255 GCGGGTATGCTGGAGCGGACTGCGAGTACCCGAGCAACCTGATATATATCTCCACCAAGT 7314  
 1080 GGGGGATTGGTTTTGAACTATGCGCAAAATCCGGATACCAACGCGCTAGCTGCGGTG 1139  
 7315 GGGGGATTGGTTTTGAACTATGCGCAAAATCCGGATACCAACGCGCTAGCTGCGGTG 7374  
 1140 AAGATGTTTACCAAAACGATGGATCAGTTGCGTAAATCCGAAAAAATTTGGATCTGAAAAAT 1199  
 7375 AAGATGTTTACCAAAACGATGGATCAGTTGCGTAAATCCGAAAAAATTTGGATCTGAAAAAT 7434  
 1200 AACCCAGCGGTATTTTCCGTTGCGATTTGCTGGTTGTGAAAAATACGGTACGATAAG 1259  
 7435 AACCCAGCGGTATTTTCCGTTGCGATTTGCTGGTTGTGAAAAATACGGTACGATAAG 7494  
 1260 CTTTTCGAGCTTGATACATTCCTCAAAAGAAAGTGGCTAAAGTAAATTTGAACCTGCTGCCAC 1319  
 7495 CTTTTCGAGCTTGATACATTCCTCAAAAGAAAGTGGCTAAAGTAAATTTGAACCTGCTGCCAC 7554  
 1320 ATCTAA 1325  
 7555 ATCTAA 7560

RESULT 3  
 AJ414149  
 LOCUS  
 DEFINITION  
 Yersinia pestis strain CO92 complete genome; segment 9/20.  
 ACCESSION  
 AJ414149.1 GI:15979570  
 VERSION  
 AJ414149.1  
 KEYWORDS  
 Yersinia pestis CO92  
 ORGANISM  
 Yersinia pestis CO92  
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 Enterobacteriaceae; Yersinia.  
 REFERENCE  
 1 (bases 1 to 193050)  
 Parkhill, J., Wren, B., Thomson, N.R., Titball, R.W., Holden, M.T.G.,  
 Prentice, M.B., Sebaihia, M., James, K.D., Churcher, C., Mungall, K.L.,  
 Baker, S., Basham, D., Bentley, S.D., Brooks, K., Cerdono-Tarraga, A.M.,  
 Chillingworth, T., Cronin, A., Davies, R.M., Davis, P., Dougan, G.,  
 Feltwell, T., Hamlin, N., Holroyd, S., Jagers, K., Leather, S.,  
 Kariyeh, A.V., Moule, S., Oyston, P.C.F., Quail, M., Rutherford, K.,  
 Simmonds, M., Skelton, J., Stevens, K., Whitehead, S. and Barrall, B.G.  
 Genome sequence of Yersinia pestis, the causative agent of plague  
 Nature 413 (6855), 523-527 (2001)  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 21470413  
 11586360  
 2 (bases 1 to 193050)  
 Parkhill, J.  
 Direct Submission  
 Submitted (04-OCT-2001) Submitted on behalf of the Yersinia  
 sequencing team, Sanger Centre, Wellcome Trust Genome Campus,  
 Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk  
 AUTHORS  
 JOURNAL  
 COMMENT  
 Details of Y. pestis sequencing at the Sanger Centre are available  
 on the world wide web.  
 (URL, [http://www.sanger.ac.uk/Projects/Y\\_pestis/](http://www.sanger.ac.uk/Projects/Y_pestis/)).

```
FEATURES
source
Location/Qualifiers
1. 193050
/organism="Versinia pestis CO92"
/mol_type="genomic DNA"
/db_xref="GI:15979573"
/translation="MRPYDFIGIGPFLNLGACLTQPLNDVCCFLIDQKAGFDWHPG
MMLESATLQTPFMSDLVTLASPELSTLNTKQQRISYFIRENPFLLRKYNYQC
QWAEQNLNRENTQVDHDEDTALRYLTGSCSRHEFLYCLLCBIDRFQPLN
PCCOFFIOI VMSGOYLHVKALQKRAITTLGSGAAEFYLLCLCEIDRFQPLN
WLRSPRELPLEVTKLTLEMSPEYVDYFHLPEKRDENLTHKPLVKGINGTILND
IFDLYTKELGALNLYTSELQKVSSSHDKGLHLVLQHEQNLQFTEGAYL
AGYHQPTTFEGIRHMDNKGRIYQVRNISDRNNTIFVQNVELHGHGVTIPDL
GMACVRSCLREITGTBYTPIEKNIAPQFSAPSLRRKB"
misc_feature
2912..2944
/gene="alca"
/note="PS00013 Prokaryotic membrane lipoprotein lipid
attachment site."
gene
4180..4761
/gene="alcB"
/note="synonym: YPO1531"
4180..4761
/gene="alcB"
/note="alcB"
/note="similar to Bordetella bronchiseptica AlcB alcB
TR:P94254 (EMBL:U61153) (201 aa) fasta scores: E(): 2e-27,
43.7% id in 174 aa, and to Streptomyces coelicolor
putative acetyltransferase sccl05.15 TR:Q9L070
(EMBL:AL163641) (184 aa) fasta scores: E(): 3e-23, 39.5%
id in 172 aa"
/codon_start=1
/transl_table=11
/product="putative siderophore biosynthetic enzyme"
protein_id="CAC90354.1"
/db_xref="GI:15979574"
/db_xref="SPTREMBL:Q8ZF22"
/translation="MMNARNIESDVTVRSGGFLRPLVLAHDVTLMDWNREYAHY
WMQGVQQVKAFYQELTQPGCVFVYQQQPAFLRYLAQHDVIADYFPAQPD
DVGMLVAPVERITGTWRVQAFITFISDPTVARIVVEPDVRNEKIHRLNKRAG
FYVQOQLTLPHKTAWLAFCTREQVRLAQKEHQ"
4758..6602
/gene="YPO1532"
/758..6602
/gene="YPO1532"
/note="similar to Bordetella bronchiseptica Alcc TR:P94255
(EMBL:U61153) (618 aa) fasta scores: E(): 0.53% id in
608 aa, and to Streptomyces coelicolor hypothetical 66.6
kDa protein sccl05.16 TR:Q9L069 (EMBL:AL163641) (595 aa)
fasta scores: E(): 0.49% id in 589 aa. Similar to
YPO1538 (46.8% identity in 596 aa overlap)"
/codon_start=1
/transl_table=11
/product="putative siderophore biosynthetic enzyme"
protein_id="CAC90355.1"
/db_xref="GI:15979575"
/db_xref="GOA:Q8ZF21"
/db_xref="SPTREMBL:Q8ZF21"
/translation="MMDLAPLSPHQSLELPHLQVLMQKNRHLHLKAISEPSHEC
LLAPQMTDHPDSGYDYQLVAAADKPNYVFRARLADHLWLDPSDLHKTVNEK
STDLDLLEIFEFKQLSISERVLPYLEETSTLYSAFKHRTGIGATLVNASFQ
IIEKMEGHPFVANNRGIRGDAQDFRSPAAASDVHLVLAHKSKAHFACIAEQ
DYAKMEGBELGAEVLAEQQLIARDQNPQYVHPQWQNKLTISIADIANQR
LVFLQGDQAYQAQOSIRTFNRSHPPQRYVYKMAISILNMGFMGLSPYVATTGIN
EWLFDVGEILOQYDFKILREVASIGFRNSYEQALTGSAYKMAAALWRENPLVY
LIQPCQLMTWALHLDQEKALPLIDLSGLSTQQLWRLSKYSPLLLCLDLY
DLMFPHGENIILVQNAVPOHIFMKDLAEELVLTNEADLPEKARRIIVMDPDMKT
LTILSDVDFGVFVLAQDQGYPERYFVADVAACIRQYQQQYFELAAKFAFSLF
TPKIKRCLNRLQNNQMLDLADPVQSQFADDLINPLVEFEVSAG"
complement(6711..7559)
/gene="YPO1533"
complement(6711..7559)
/note="similar to Vibrio cholerae ferric vibriobactin ABC
transporter, ATP-binding protein vc0779 TR:Q9KTV8
(EMBL:AS004163) (279 aa) fasta scores: E(): 0.55% id in
263 aa, and to Streptomyces coelicolor putative
iron-siderophore uptake system ATP-binding component
```

sc151.25C TR:Q9S215 (EMBL:AL109848) (301 aa) fasta scores:	
E(): 0, 57.2% id in 264 aa"	
/codon_start=1	
/transl_table=11	
/product="putative iron-siderophore transport system,	
ATP-binding component"	
/protein_id="CAC90356.1"	
/db_xref="GI:15979576"	
/db_xref="GOA:Q9ZF20"	
/db_xref="SPTRMBL:Q8ZF20"	
/translation="MRSQRILTFPMLPENGDISDKPLRTTQKSLGSGQOOLINKL	
SLSKNGAFSVIIPNGCGKSTLRSLRSLIPQNGSLRDQDITQHYKAKVFARQLS	
LSLGSAGSEITLVYDNGRVAHVPHOWSTEDERIVKAAISAVNLSLIVQORVS	
ELSGQRCAFWATFLAQHPRIVLDEPTTYLDIAHOIEMLDLCOELQOQGTALVLVL	
HDINQALRYATHLLIMKEGQIYAEGLPESIVIEDSIAAVFGILRCHLIIDPSGKPLVI	
PRRQSR"	
complement(6849..7409)	
/gene="YPO1533"	
/note="Pfam match to entry PF00005 ABC tran, ABC	
transporter, score 183.50, E-value 3.5e-51"	
complement(7032..7076)	
/gene="YPO1533"	
/note="PS00211 ABC transporters family signature."	
Query Match	98.6%; Score 1306; DB 1; Length 193050;
Best Local Similarity	99.5%; Pred. No. 0;
Matches 1320; Conservative	0; Mismatches 5; Indels 1; Gaps 1;
QY 1	ATGTCAGTATTAGAAAATCGTGACGCTACTCTGGCTGCTATGATGCTAGCCGATG 60
Db 135753	ATGTCAGTATTAGAAAAATCGTGACGCTACTCTGGCTGCTATGATGCTAGCCGATG 135812
QY 61	GCTGCTATTACTGCGCCCGGTAGCCGCCAGCCATCGGGCTATACTTTAGAACGTGTGTT 120
Db 135813	GCTGCTATTACTGCGCCGTAGCCGCCAGCCATCGGGCTATACTTTAGAACGTGTGTT 135872
QY 121	ATTTTGAGTCGCATGTGTTCGTCGCCGACCAAAC-AACGCAGCTTATGATGATGTT 179
Db 135873	ATTTTGAGTCGCATGTGTTCGTCGCCGACCAAACAGCAGCTTATGATGATGTT 135932
QY 180	AGCCAGATATAGTGCCACCATTGGCCGGTAAAGCGGGGTATTTAACGCCACGTGTGCG 239
Db 135933	AGCCAGATATAGTGCCACCATTGGCCGGTAAAGCGGGGTATTTAACGCCACGTGTGCG 135992
QY 240	GAGTTGGTCACATTCATGGGGGGGTTTATCGTGATTACTTTTCGACGCTTGGTTGTTA 299
Db 135993	GAGTTGGTCACATTCATGGGGGGGTTTATCGTGATTACTTTTCGACGCTTGGTTGTTA 136052
QY 300	GGGCGGGATGTCGCGCAGAGGGGGGTATATGCACAGCAGATATCATCAACGTACC 359
Db 136053	GCGCGGGATGTCGCGCAGAGGGGGGTATATGCACAGCAGATATCATCAACGTACC 136112
QY 360	CGCTTAACCGGACAGGCATTTCTTGATGCTGTGCTCCGGGTGTGTTTGACCCGTGCAT 419
Db 136113	CGCTTAACCGGACAGGCATTTCTTGATGCTGTGCTCCGGGTGTGTTTGACCCGTGCAT 136172
QY 420	AATCAGCGCATTTGAAAAGACCATCCCTGTTCCATCCGCTAGAGATCGCGGTGTGT 479
Db 136173	AATCAGCGCATTTGAAAAGACCATCCCTGTTCCATCCGCTAGAGATCGCGGTGTGT 136232
QY 480	AAGTTAGACAAACGCACAAACAGATTAAGCGATTCGAAGAACGATTTGGCGCGCGTTAGAT 539
Db 136233	AAGTTAGACCGCGCACAAACAGATTAAGCGATTCGAAGAACGATTTGGCGCGCGTTAGAT 136292
QY 540	ACGGTAGCAGCGCTACGCTAAAACGTTTGGCCAGATGGGGGACGTGCTGAATTTGCG 599
Db 136293	ACGGTAGCAGCGCTACGCTAAAACGTTTGGCCAGATGGGGGACGTGCTGAATTTGCG 136352
QY 600	GCCTTCCTCTATTGCAAAATCTTTGCAACAGCAAGGAAAAACGTGTGATTTTGCCACTTT 659
Db 136353	GCCTTCCTCTATTGCAAAATCTTTGCAACAGCAAGGAAAAACGTGTGATTTTGCCACTTT 136412
QY 660	GGCGCCAAATGAGTTTAACGTTTAATAAGAAGGGACAAAAGTGAACCCCTCAGTGGGCCACTG 719

Matches 1316; Conservative 0; Mismatches 8; Indels 2; Gaps 2;  
 QY 1 ATGTCAGTATTAGAAATCGGTGACGGCTATCTCGGCTGTTATGATGCTTAAGCGGATTC 60  
 DB 1 ATGTCAGTATTAGAAATCGGTGACGGCTATCTCGGCTGTTATGATGCTTAAGCGGATTC 60  
 QY 61 GCTGCTATTACTGCGCGGTAGCGCGGAGCCATCGGCTATACCTTTAGAAACGTTGGTT 120  
 DB 61 GCTGCTATTACTGCGCGGTAGCGCGGAGCCATCGGCTATACCTTTAGAAACGTTGGTT 120  
 QY 121 ATTTTGGTCCGATGTTGTTGCTCGCGACCAAC- AACGACGCTTATGATGATTT 179  
 DB 121 ATTTTGGTCCGATGTTGTTGCTCGCGACCAAC- AACGACGCTTATGATGATTT 179  
 QY 180 ACGCCAGATAAGTGGCCCAATAGCGCGGTAAAGCGGGTATTTAAACGCCACGTTGGCG 239  
 DB 180 ACGCCAGATAAGTGGCCCAATAGCGCGGTAAAGCGGGTATTTAAACGCCACGTTGGCG 240  
 QY 240 GAGTTGGTCAATGATGGGGGGTATTTATGGTATTAATCTTCGACGCTTGGTTGTTA 299  
 DB 240 GAGTTGGTCAATGATGGGGGGTATTTATGGTATTAATCTTCGACGCTTGGTTGTTA 299  
 QY 300 CGCGCGGATGTCGCGCAGAGGGGGGTATATGACAGGAGATATCGATCAACGATACC 359  
 DB 300 CGCGCGGATGTCGCGCAGAGGGGGGTATATGACAGGAGATATCGATCAACGATACC 360  
 QY 360 CGTTTACCGGACAGGCAATTTCTGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 419  
 DB 360 CGTTTACCGGACAGGCAATTTCTGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 420  
 QY 420 AATCAGGCGGATTTGAAAGACCGATCCCTGTTTCCATCGGTAGAGACTGGGCTGTT 479  
 DB 420 AATCAGGCGGATTTGAAAGACCGATCCCTGTTTCCATCGGTAGAGACTGGGCTGTT 480  
 QY 480 AAGTTAGACACGCAACAGATAAGCGATTAAGAAACGATTTGGCGGGCGCTTATGAT 539  
 DB 480 AAGTTAGACACGCAACAGATAAGCGATTAAGAAACGATTTGGCGGGCGCTTATGAT 540  
 QY 540 ACGGTGACCGGCTACGCTAAACCTTTTCCAGAGTGGGGAGCTGCTCAATTTTGGC 599  
 DB 540 ACGGTGACCGGCTACGCTAAACCTTTTCCAGAGTGGGGAGCTGCTCAATTTTGGC 600  
 QY 600 GCTTCTCTTATGCAATCTTTGCAACAGCAAGGAAACGTTGATTTTGGCCACTTT 659  
 DB 600 GCTTCTCTTATGCAATCTTTGCAACAGCAAGGAAACGTTGATTTTGGCCACTTT 659  
 QY 660 GCGGCAATGAGTTAAAGTTAAAGAGGACAAAGTGAACCTCAGTGGGCGCACTG 719  
 DB 660 GCGGCAATGAGTTAAAGTTAAAGAGGACAAAGTGAACCTCAGTGGGCGCACTG 719  
 QY 720 GCGTTATCATCGACATTTGGTGAATCTTTTATTAACAAACGCAAGCCATGCCAGAG 779  
 DB 720 GCGTTATCATCGACATTTGGTGAATCTTTTATTAACAAACGCAAGCCATGCCAGAG 779  
 QY 780 GTTGTGCGCAAGGCTAAAGGGGCGGAGAAATGGGTATCTTTATTTGCTATTAACATAC 839  
 DB 780 GTTGTGCGCAAGGCTAAAGGGGCGGAGAAATGGGTATCTTTATTTGCTATTAACATAC 839  
 QY 840 GCGCAATTTAATTTGCAAAACACCGTATATCGCCCGTCAATAAGGAGCGCCATTA 899  
 DB 840 GCGCAATTTAATTTGCAAAACACCGTATATCGCCCGTCAATAAGGAGCGCCATTA 899  
 QY 900 TTACAGCAGATAGATACGCTTTAAACCTTTGATGCTCAGGGGCAAAAGTACCC 959  
 DB 900 TTACAGCAGATAGATACGCTTTAAACCTTTGATGCTCAGGGGCAAAAGTACCC 959  
 QY 960 ATTTTACGCCCCAAACCGGGTCTTTGCTCGGTGGGCGATGATACCAATATTGCCAATATT 1019  
 DB 960 ATTTTACGCCCCAAACCGGGTCTTTGCTCGGTGGGCGATGATACCAATATTGCCAATATT 1019  
 QY 1020 GCGGGTATGCTGGGAGCGGCTACCGAGCTTACCGAGCAACCTGATTAATCTCCACGAGT 1079  
 DB 1020 GCGGGTATGCTGGGAGCGGCTACCGAGCTTACCGAGCAACCTGATTAATCTCCACGAGT 1079

QY 1080 GGGGATTTGGTTTGAACATATGGCAAAATCCGATAACCAACGCGCTACGTTCCGGTG 1139  
 DB 1080 GGGGATTTGGTTTGAACATATGGCAAAATCCGATAACCAACGCGCTACGTTCCGGTG 1139  
 QY 1140 AAGATGTTTACCAACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1199  
 DB 1140 AAGATGTTTACCAACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1199  
 QY 1200 AACCACGCGGTATTATTATTTCCGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1259  
 DB 1200 AACCACGCGGTATTATTATTTCCGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1259  
 QY 1260 CTTTGGAGCTGATACATTCCTCAAAAGAAAGTGGCTAGGTTAATTTGAACCTGCTGCCAC 1319  
 DB 1260 CTTTGGAGCTGATACATTCCTCAAAAGAAAGTGGCTAGGTTAATTTGAACCTGCTGCCAC 1319  
 QY 1320 ATCTAA 1325  
 DB 1320 ATCTAA 1325

RESULT 5  
 AX556799  
 LOCUS 1326 bp DNA linear PAT 27-NOV-2002  
 DEFINITION Sequence 11 from Patent WO0248332.  
 ACCESSION AX556799  
 VERSION AX556799.1 GI:25899907  
 KEYWORDS Yersinia pestis  
 SOURCE Yersinia pestis  
 ORGANISM Yersinia pestis  
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 Enterobacteriaceae; Yersinia.  
 REFERENCE 1  
 AUTHORS Short, J., Mathur, E.J., Richardson, T., Robertson, D. and Barton, N.  
 TITLE Recombinant phycocyanins and uses thereof  
 JOURNAL Patent: WO 0248332-A 11-20-JUN-2002;  
 DIVERSA CORPORATION (US)  
 FEATURES  
 Location/Qualifiers  
 1..1326  
 /organism="Yersinia pestis"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:632"

ORIGIN  
 Query Match 62.5%; Score 827.6; DB 6; Length 1326;  
 Best Local Similarity 77.0%; Pred. No. 6e-237;  
 Matches 1021; Conservative 0; Mismatches 304; Indels 1; Gaps 1;  
 QY 1 ATGTCAGTATTAGAAATCGGTGACGGCTATCTCGGCTGTTATGATGCTTAAGCGGATTC 60  
 DB 1 ATGTCAGTATTAGAAATCGGTGACGGCTATCTCGGCTGTTATGATGCTTAAGCGGATTC 60  
 QY 61 GCTGCTATTACTGCGCGGTAGCGCGGAGCCATCGGCTATACCTTTAGAAACGTTGGTT 120  
 DB 61 GCTGCTATTACTGCGCGGTAGCGCGGAGCCATCGGCTATACCTTTAGAAACGTTGGTT 120  
 QY 121 ATTTTGGTCCGATGTTGTTGCTCGCGACCAAC- ACGCAGCTTATGATGATTT 179  
 DB 121 ATTTTGGTCCGATGTTGTTGCTCGCGACCAAC- ACGCAGCTTATGATGATTT 179  
 QY 180 ACGCCAGATAAGTGGCCCAATAGCGCGGTAAAGCGGGTATTTAAACGCCACGTTGGCG 239  
 DB 180 ACGCCAGATAAGTGGCCCAATAGCGCGGTAAAGCGGGTATTTAAACGCCACGTTGGCG 240  
 QY 240 GAGTTGGTCAATGATGGGGGGTATTTATGGTATTAATCTTCGACGCTTGGTTGTTA 299  
 DB 240 GAGTTGGTCAATGATGGGGGGTATTTATGGTATTAATCTTCGACGCTTGGTTGTTA 299  
 QY 300 CGCGCGGATGTCGCGCAGAGGGGGGTATATGACAGGAGATATCGATCAACGATACC 359  
 DB 300 CGCGCGGATGTCGCGCAGAGGGGGGTATATGACAGGAGATATCGATCAACGATACC 360

360	CGCTTAA	CCGGACAG	CAGCATTT	CTTGATG	TGTGGCT	CCGGGGT	TGTGGT	TGTGAC	CGGTGCAT	419	
361	CGTTTAA	CCGGTCAG	CGCTTTT	CTGGAAT	GTTGGCG	CGGGT	TGGCGCT	TGACTGT	CTCCAC	420	
420	AATCAG	CGCGATT	TGAAAAG	ACCGAT	CCCTCT	TTCAT	CCCGTAG	ACTGCG	CTGT	479	
421	AATCAG	CGCGATT	TGAAAAC	CCGAT	CTCTGT	TTCAT	CCCGT	TGAAA	ACCGCGT	480	
480	AAATTAG	ACAA	CGCACA	CAGATA	TAAG	CGATT	TGAAGA	ACGATT	TGGCG	CGCGT	539
481	AAACTG	GAACA	ACGCCCAA	CCGATA	AGGCA	ATT	TGAGAA	CGCT	TGGCG	CGCGT	540
540	ACGGT	CAGC	CGCGCT	TACG	TAAAC	CTTT	TGCCAG	ATGGGG	ACGCT	TGTAAT	599
541	ACGGT	AAAG	CCAGCG	CTAT	TGCCAA	ACCGT	TTCGCA	ANTGG	CGAT	TCTCT	600
600	GCTTCT	CCCTTA	TGCGAA	TCTTT	TGCAAC	GACGAA	GGAAAA	CGTGT	GAATTT	TGGCC	659
601	GCGAGT	CCGTACT	CGCAAGT	CAC	TGCA	GACGAG	GGGAAAA	CTTGT	GACTT	TCCAC	660
660	GCGGCC	CAATGA	AGTTAA	CGTTA	TAAAG	GAGGAC	AAAA	GTGAC	CCCT	CAGTGG	719
661	GCGGCC	CAACGA	AGTTA	TGTAA	CAGAA	GAGGAC	CGAA	AGTTAC	CTCT	TGAC	720
720	GCGTTAT	ATCG	GATTTGG	TGGTG	AAATCT	CTTTA	TATTA	CAAAA	CGCA	ACGCA	779
721	GCGCTG	CTAG	CA	CGTTGG	CGAAATCT	CTTCT	TGCTG	CA	AGC	GACG	780
780	GTTGCT	CGCA	ACGGCT	TAAAA	GGCGG	GAGAA	TTGG	TAT	CTTAT	TGT	839
781	GTAG	CGTGG	CAGCGT	TTGAA	GGCGT	TGAG	ACTGG	TGCT	CTCT	TGAG	840
840	GCGCAAT	TTAA	TTAAT	TGCA	AAAA	CA	CGTAT	TATCG	CCCGT	CA	899
841	GCA	CAGTTCA	ACCTG	ATGGT	TAAAG	CGCCAT	CA	ATTG	CA	CGCC	900
900	TTAC	GACAG	ATAG	ATAC	GGCTTT	TAAC	CTTCA	CTG	ATGCT	CAG	959
901	TTAC	GCAAT	CGAT	ACCG	CACT	TGAC	CTCT	GCA	CTG	GAC	960
960	ATTTAC	CGCCAA	AAACCG	GGTCT	TGTC	TG	TGGG	CGAT	TGAT	ACCA	1019
961	ATCT	CGGCT	CAGAA	CCGTGTTT	TATCT	CTGG	TGGC	ACG	ACAC	AAAT	1020
1020	GCGGGT	ATGCT	GGAG	CCGACT	CGG	AGCTAC	CCG	GAGCA	CCCTG	ATAAT	1079
1021	GCGCGT	ATGCT	GGGCG	AGATTG	CGAGT	TATAC	CG	AAAC	CCCG	AT	1080
1080	GCGGGAT	TGTT	TGA	ACTAT	CGCA	AAAT	TC	CGGATA	CCAC	CAG	1139
1081	GCGCGT	CTGCT	TTTG	AGCTGT	TGGC	AGAA	TC	CGCA	CAAT	CAT	1140
1140	AGAT	GTCTTAC	CAAA	CGAT	TGAT	CGGT	TAAT	GCCG	AAAA	TTGG	1199
1141	AA	GATGTT	CTAT	CAG	ACCAT	TGGAT	CAAT	TGCG	TAA	CGCG	1200
1200	AA	CCCG	CGCG	GTAT	TTT	CCGT	TG	CAGT	TGCT	GGTT	1259
1201	AA	TC	CCCG	CG	GCAT	CAT	CAGT	GT	CGCT	TG	1260
1260	CTTT	TG	CG	AGCT	CTG	AT	CTT	CA	AA	GA	1319
1261	CTG	TG	GA	ACT	CT	GAT	CTTT	CA	AAAA	AA	1320
1320	AT	CTAA									1325
1321	ATT	TAA									1326

RESULT	6
AY1378096	
LOCUS	AY1378096 2558 bp DNA linear BCT 06-OCT-2003
DEFINITION	Obesumbacterium proteus phytase gene, complete cds.

ACCESSION	REFERENCE
VERSION	AUTHORS
KEYWORDS	JOURNAL
SOURCE	REFERENCE
ORGANISM	AUTHORS
	TITLE
	JOURNAL

REFERENCE AUTHORS JOURNAL	REFERENCE AUTHORS TITLE JOURNAL
1. H. J. ...	2. ...
3. ...	4. ...
5. ...	6. ...
7. ...	8. ...
9. ...	10. ...
11. ...	12. ...
13. ...	14. ...
15. ...	16. ...
17. ...	18. ...
19. ...	20. ...
21. ...	22. ...
23. ...	24. ...
25. ...	26. ...
27. ...	28. ...
29. ...	30. ...
31. ...	32. ...
33. ...	34. ...
35. ...	36. ...
37. ...	38. ...
39. ...	40. ...
41. ...	42. ...
43. ...	44. ...
45. ...	46. ...
47. ...	48. ...
49. ...	50. ...
51. ...	52. ...
53. ...	54. ...
55. ...	56. ...
57. ...	58. ...
59. ...	60. ...
61. ...	62. ...
63. ...	64. ...
65. ...	66. ...
67. ...	68. ...
69. ...	70. ...
71. ...	72. ...
73. ...	74. ...
75. ...	76. ...
77. ...	78. ...
79. ...	80. ...
81. ...	82. ...
83. ...	84. ...
85. ...	86. ...
87. ...	88. ...
89. ...	90. ...
91. ...	92. ...
93. ...	94. ...
95. ...	96. ...
97. ...	98. ...
99. ...	100. ...

REFERENCE AUTHORS	TITLE JOURNAL	REFERENCE AUTHORS	TITLE JOURNAL
1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16
17	18	19	20
21	22	23	24
25	26	27	28
29	30	31	32
33	34	35	36
37	38	39	40
41	42	43	44
45	46	47	48
49	50	51	52
53	54	55	56
57	58	59	60
61	62	63	64
65	66	67	68
69	70	71	72
73	74	75	76
77	78	79	80
81	82	83	84
85	86	87	88
89	90	91	92
93	94	95	96
97	98	99	100

FEATURES SOURCE

CDS

## ORIGIN

Query Match 24.4%; Score 323; DB 1; Length 2558;  
Best Local Similarity 57.8%; Pred. No. 3.1e-85;  
Matches 633; Conservative 0; Mismatches 490; Indels 16; Gaps 6;

Qy	86	CCGAGCATTGGGCTATCTACTTTAGACGTTGGTTATTTTGGTTCGCTACGTCACATGGTGTTCGGCT	145
Db	1067	CCGACCTTCCGGGATATCAGCTTGAAGAAGTGTTGTTCTCACCOCGGCATGGCTTTCGAG	1126
Qy	146	CGCCGACCAAAACA - ACGAGCTTATGAATGATTTACGCCAGATTAAGTGCACCAATGGC	204
Db	1127	CGCCAAACCAAAATGACGCAAAACCATCGCGAGTAAACCTTAATGCGTGCCTTGATGTGC	1186
Qy	205	CGGTAAAGCGGGTATTTTAAAGCCACAGTGGTGCAGATTGGTTCACATTGATGGGGGGGT	264
Db	1187	CAGTCAAACTGGGGTATATCAGCCAGAGGGCGAGCATTTGGTGAAGCTTGATGGGTGGGT	1246
Qy	265	TTTATCGTGATTACTTTTCGACGCTTGG --- TTTGTTAGCGCGGATGTCGCGCAGAGG	321
Db	1247	TC'TATCGGAGAGTATTTCAGCAGTTGGGGATCTTGTCTAAAGTTCGCTGTCCAACGGCGCA	1306
Qy	322	GGGGGGTATATGCACAGGCAGATATCGATCAACGTAACCGCTTAAACGGACAGGCATTTC	381
Db	1307	ATCATGTTTATGTTTGGGACAGCGTTCGACCAAGCCACTCGGAAAACAGGCGAGCGCTTTT	1366

Phytases  
(in) 2ND MOSCOW INTERNATIONAL CONGRESS "BIOTECHNOLOGY: STATE OF THE  
ART AND PROSPECTS OF DEVELOPMENT";  
Nova Science Publishers, Inc., New York, NY, USA)

3 (pages 1 to 2558)

Zinin, N. V., Samsonov, V. V., Borshchevskaya, L. N., Kanikovskaya, A. A., Gudima, M. Y., Syneokii, S. P. and Syneokii, S. P.

Phytase activity of several bacteria groups  
Biotechnologia (Mosc.) 13, 3-10 (2003)

4 (bases 1 to 2558)  
Zinin, N.V. and Syneok, S.P.

Direct Submission  
Submitted (02-SEP-2003) VKPM, Gos NII Genetika, 1st Dorozny Proezd,

**Moscow 117545, Russia**  
**Location/Qualifiers**

1. .2558  
/organism="Obesumbacterium proteus"

```
/mol_type="genomic DNA"  
/db_xref="taxon:82983"
```

961. .2295  
/note="PhyA"

```

/codon_start=1
/transl_table=11
--3--"

```

```

/product="phytase"
/protein_id="AAQ90419.1"
/submitter="NCBI"
/submitter_email="ncbi-ncbi@ncbi.nlm.nih.gov"

```

```

/db_xref="GI:37266291"
/translation="MTISLFTSPTRLKCMPLAFIAASMLTTASYASETEPSGYQLE
VWVLI SPCVSA QMTATQMTQNTDVAWDEWVYI CVITDPGEHI VSI MCGFEYRKQFO

```

QLGILSGRCPTANDVYVWADVDRTRKTGEAFLAGLAPAECHLSIHHQDPLF  
 HTVAVACUETKENTONVAVEQACMBPITDOI NOHYBPALAI MSSV NEPKSTYCOOHS  
 KWILSRHGVRAPTKMTQIMRDUITPNAWPEWPKLGIITPRKGHLVSLTGGGFIKQNF

QDQCDLAQAMPKSLSIKDNQNKVALDGAVGLSSITLAEIIFLLEYAQGMPDEAAWGKIHSE  
OCDQNAIIFTHNAQENI MSPBTVI AVHNCBTEI ILOTTIVSATINSOPSSREI BELSADNKIL  
HPVKVAGCIMEKTVQQAQVEQAQGFIDQLNQLHCFKFAHAISSVAVNF FASGICQGHSE

QWNNALLTTHNAQFDEMSKIPYIAHHNGIFLEQJIVLSAINSQFSKRELFEBQADAKAL  
PPAGHDNTIANIAGMFGSWALPGQPDNTPPGALVFERMSDKTKYKYSVSVQMYQTL  
BOI.PNOTRPI.TDKPAGSVAIKPGCDDOTAEVCYPI.DTETRLAKONELVECO"

AQLRNQIFLTIILDRFAGSVALLKIFGCDDQIABGICFDDIFIKRANQNEIVECQ

ch 24.4%; Score 323; DB 1; Length 2558;  
57.8%; Pred. No. 3.1e-85;  
Similarity

693; Conservative 0; Mismatches 490; Indels 16; Gaps 6;  
1 Similarly 37.8%; Fred: NO. J.1.E.03;

86 CCGAGCCATCGGCTATACCTTAGAACGTGTGGTTATTTGAGTCGCCATGGTGTTCGCT 145

67 CCGAACCTTCGGGATATCAGCTTGAAAGGTGGTGATTCTCAGCCGGCATGGCGTTCGAG 1126

46 CGCCGACCAACA - ACGCAGCTTATGAATGATGTTACGCCAGATAAGTGGCCACAATGGC 204

27 CGCCAACCAAATGACGGCAACCATGGCGACGTAAACCTAATGCGTGGCCTGAGTGGC 1186

05 CGGTAAAGCGGGTATTTAA CGGCCA CGTGGT GCGGAGT TGGTCA CATTGATGGGGGGT 264

87 CAGTCAAACCTGGGGTATATCAGCCCGAGGGCGAGCATTTTGGTGAGCTGATGGGTGGGT 1248

65 TTTATGGTGATTACCTTCGCAGCCITGG---TTTGTTAGCGGGCGGATGTCGGGCAGAGG 321

47 TCTATCGGCGAAGTTTCAGCAGTTGGGGATCTTGTCTAAAGGTCGGTGTCTCAACGGCGA 1308

222 GGGGGGTATATGCACAGGCAGATATCGATCAACGTTACCCGCTTAAACGGACAGGCATTTC 381

07 ATGATGTTTATGTTTGGGCAGACGTCGACCAACGCCTCGGAATAACAGGCCGAGGCCGTTT 136



QY 382 TTGATGTGTGGCTCCGGGTGTGGTTTGACCGTGCAATAATCAGCCGATTTGAAAAAGA 441  
Db 1367 TAGCGGTCTCGCACCTGAATGCCATCTAAGTATTCATCATCAGCAAGACATCAGCAGG 1426  
QY 442 CCGATCCCTGTTTCATCCGCTAGAGACTGGCGTGTGTAAAGTTAGACAACGCAACAG 501  
Db 1427 CCGACCCGCTATTTTCATCCGCTGAAGCGGTGTCTGCAATGGAAGAAACGCGGTAC 1486  
QY 502 ATAAGCGATTGAAGAACGATTGGCGGGCGGTGTAGATAGCAGTAAACAGCAGCTACGCTA 561  
Db 1487 AGCAGCGCTGAGCAACAGCGCGGTATGCCATAGATCAGTAAACAGCAGCTACGCGC 1546  
QY 562 AACCTTTTCCAGATGGGAGACGCTGCTGAATTTTGGCGCTTCTCCTTATGCAATCTT 621  
Db 1547 CTGCACTGGCGCTGATGAGCAGCGTGTGCTGCAATTTCCAAAGTCAACGTACTGCCA---GC 1603  
QY 622 TGCAACAGCAAGGAAACGCTGATTTTGCCCACTTTGGCGGCAATGAAGTTAAAGTTA 681  
Db 1604 AACATAGCGCAGCAACCACTGCGATTAGCAAGCGCATGCCAGCAAGCTCTCAATCA 1663  
QY 682 ATAAGAGGAGCAAAAGTAGTACCCTCAGTGGGCCACTGGCGTTATATCATGACATTTGGGT 741  
Db 1664 AAGCAATGCAATAAAGTGCACTTGTATGGGCGGTAGTGTGCTCAACGCTGGCGG 1723  
QY 742 AAATCTTCTTATTACAAAACGCAAGCCATGCCAGAGGTGCTGCGCAACGCTTAAAG 801  
Db 1724 AAATTTCTCTTGAATATGCGCAGGCGATGCCAGATGCGCGTGGGGAATAATCCATT 1783  
QY 802 GGGCGGAGAAATGGGTATCTTTATGTCATTACATAACGCGCAATTTTAAATTTGACAA 861  
Db 1784 CAGAGCAAGACTGGAACGCTTTGCTAACGCTTACAAACGCAATTTGATGTGTAC 1843  
QY 862 AAACACCGTATATCGCCGCTATTAAGGAGCGCAATTTATACAGCATAGATACGCTT 921  
Db 1844 GCACGCTTATACGCAACATTAACGCGCACCGCTATTACAAACCATCGTCAGCGC-- 1901  
QY 922 TAACCTTTCAACTGGATGCTCAGGGGCAAAAGCTTACCCATTTCAGCCCAAAACCGGGTCT 981  
Db 1902 -AATCAATTCACACCTAGCTCGGAGATTGCGCGAGTTATCAGCGATACAAAGATCC 1960  
QY 982 TGTTCTCGTGGCATGATACCAATATGCAATATTTGGCGGTATGCTGGAGCCGACT 1041  
Db 1961 TGTTCCGCGCGCATGACACCAATATGCAATATGCGGGGTATGTTGGGTATGTCGT 2020  
QY 1042 GGCAGCTACCGGAGCAACTGATTAATCTCCACAGGTGGGGATTGGTTTTTGAACAT 1101  
Db 2021 GGGCGCTACCGGCGCGCGATTAACAGCTCCAGCGGTGCTTATGTTTGAAGCTT 2080  
QY 1102 GGCAAAATCCGGATAACACACGCGCTACGTTGCGGTGAAGATTTCTACCAACGATGG 1161  
Db 2081 GGTGAGAT---AAGACGGGGGAAAAATACGTGAGGTACAGATGATGTACCAACGCTGG 2137  
QY 1162 ATCAGTTGCGTAATGCCGAAAAATTGGATCTGAAATATACCCAGCGGTATTTTCCG 1221  
Db 2138 CACAGCTACGTAATCAACCCACTCTCTG---GATAAGCCGCTGGCAGCGGTGGCG 2194  
QY 1222 TTGAGTTGCTGGTTGTGAAAAATAACCGGTGACGATAAGCTTTGCGAGCTTGATACATTC 1280  
Db 2195 TAAAAATACCGGGGTGTGACGATCAACCGCGGAAGGATATGGCCACTAGATACCTTC 2253

RESULT 7  
AY425982  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE

AY425982  
Obesumbacterium proteus PhylA2 gene, complete cds.  
AY425982  
AY425982.1 GI:37594476  
Obesumbacterium proteus  
Obesumbacterium proteus  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Obesumbacterium.  
1 (bases 1 to 2161)

AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source  
CDS  
ORIGIN

Zinin,N.V., Syneoky,S.P. and Debabov,V.G.  
Obesumbacterium proteus phytase sequence  
Patent: RU (2003127890)-A 17-SEP-2003;  
VKPM, GosNII Genetika; 1-st Dorozny Proezd; Moscow;  
Russia;  
2 (bases 1 to 2161)  
Zinin,N.V. and Syneoky,S.P.  
Direct Submission  
Submitted (01-OCT-2003) VKPM, GosNII Genetika, 1-st Dorozny Proezd,  
Moscow 117545, Russia  
Location/Qualifiers  
1. 2161  
/organism="Obesumbacterium proteus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:82983"  
476..1810  
/note="phytase; pH 1.5-6.5"  
/codon\_start=1  
/transl\_table=11  
/product="PhyA2"  
/protein\_id="AAQ94187.1"  
/db\_xref="GI:37594477"  
translation="MTIYLFTHSPARKLRLALIAASMLTTASYASETPSPGYOLE  
KVIILSRHVRAPMTMTCTMDVTPPEVVKLYITPRGHLASLMGPFYRQKFL  
QGLILSKDRCPANDVYTAADVDQRTKTEAFLAGLAPECHLSFHQODIKQVDSLF  
HPVKGVCITKEKAQQVQEQQAQQLQKHQRPALMSVNLNFTSYACQHSQA  
EQTCTFALVMPKSLISKDDGNKVALDGAIVLSSTLAEIIFLLEHQAQMPNAWKGHSE  
QDMNTLLTHNAQFDMISRTPIAKNGTPLLQITVSAINSQTPRELPESADNKKIL  
FLAGHDTNLANIAGMLSLPQDNTPGGALVFERWSDTKTKTVSAQMYQTLL  
AQLRNTPLTLDEPAGSVALKIPGCDQTAEGYCPDLDTFTRLAKQNELAEQ"

Query Match 24.3%; Score 321.4; DB 1; Length 2161;  
Best Local Similarity 57.7%; Pred. No. 9.3e-85;  
Matches 692; Conservative 0; Mismatches 491; Indels 16; Gaps 6;  
QY 86 CCGAGCCATCGGGCTATCTTTAGAACGTTGTTGTTTATTTGAGTCGCCATGTTGCTGT 145  
Db 582 CGTGCTTCAGCTATCAGCTTGAAGGTGTGATCTCAGCGCTCATGGCGTTTCAG 641  
QY 146 CGCGGACCAACA-ACGCGCTTATGATGATGTTACGCCAGATAGTGGCCACATGCG 204  
Db 642 CGCGACCAAAATGACGCAACCATGCGGTGTTTACGCTTAATACCTGGCGCTGAATGCG 701  
QY 205 CGGTAAACGCGGGTATTTAAACGCGGTTGGGAGTGGTTCATTTGATGGGGGT 264  
Db 702 CTGTTAAGCTGGGTATATCAACCGAGGCGCAATTTAGCGAGCTGATGGGGGAT 761  
QY 265 TTTATGTTGATTACTTTCAGCGCTTGG---TTTGTAGCGCGGGATGTTCGGCAGAGG 321  
Db 762 TCTATCGACAGAAATTTTCAGCGAGTTGGGATCCTGTCTAAAGATCGCTGCCCAACGCGCAA 821  
QY 322 GGGGGTATATGCAAGCGAGATATCGATCAAGTACCGCTTACCGGACAGCGATTTTC 381  
Db 822 ATGATGCTATTTTGGGCGAGCTGCTGATCAGCGCACCGGAAACAGCGAGGGGCTTTT 881  
QY 382 TTGATGTTGCTGCTCGGGGTGTGTTTGACCGTGCATAATCAGGCCGATTTTGAAGAAAGA 441  
Db 882 TAGCGGTCTCGCACCTGAATGCCATCTAAGTATCCATCATCAGCAAGACATCAAGCAGG 941  
QY 442 CCGATCCCTGTTCCATCCGCTAGAGACTGGCGTGTGTAAAGTTAGACAACGCAACAG 501  
Db 942 TCGATTGCTATTTTCATCCAGTGAAGCGGTGTCTGCAACCATGGAAGAGCGCGGTAC 1001  
QY 502 ATAAAGCGATTGAAGAACGATTGGCGGGCGCTTATAGATACGCTGAGCCAGCGCTACGCTA 561  
Db 1002 AGCAGCGCTCGAGCAGCAGCGCGGTATGCCCATCGCTCAGTTAAACCAACATCCCGCC 1061  
QY 562 AACCTTTTCCAGATGGGAGCGTGTGAAATTTTGGCGCTTCTCCTTATTTGCAATCTT 621  
Db 1062 CTGCACTGGCGGTGATGAGCAACGCTTAACTTCCCAACGCTCAGCGTATTGCCCA---GC 1118  
QY 622 TGCAACAGCAAGGAAAAACGTTGATTTTGGCCCACTTTTGGCGGCAATGAAGTTAAGTTA 681





LOCUS ECOACPHD 1901 bp DNA linear BCT 04-APR-2002  
 DEFINITION Escherichia coli periplasmic phosphoanhydride phosphohydrolase  
 (appa) gene, complete cds.  
 ACCESSION L03373  
 VERSION L03373.1 GI:145182  
 KEYWORDS acid phosphatase; periplasmic phosphoanhydride phosphohydrolase.  
 SOURCE Escherichia coli  
 ORGANISM Escherichia coli  
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 Enterobacteriaceae; Escherichia.  
 REFERENCE 1 (sites)  
 Dassa, J., Marck, C. and Boquet, P.L.  
 The complete nucleotide sequence of the Escherichia coli gene appA  
 reveals significant homology between pH 2.5 acid phosphatase and  
 glucose-1-phosphatase  
 J. Bacteriol. 172 (9), 5497-5500 (1990)  
 JOURNAL 90368616  
 MEDLINE 2168385  
 PUBMED  
 REFERENCE 2 (bases 1 to 1901)  
 Ostranin, K., Harms, E.H., Stevis, P.E., Kuciel, R., Zhou, M.M. and Van  
 Eeten, R.L.  
 Overexpression, site-directed mutagenesis, and mechanism of  
 Escherichia coli acid phosphatase  
 J. Biol. Chem. 267 (32), 22830-22836 (1992)  
 JOURNAL 93054596  
 MEDLINE  
 PUBMED 1429631  
 FEATURES  
 source  
 /organism="Escherichia coli"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:562"  
 188..1486  
 /gene="appa"  
 188..1486  
 /gene="appa"  
 /standard\_name="acid phosphatase"  
 /citation=[1]  
 /codon\_start=1  
 /translation="translating frame 1"  
 /product="periplasmic phosphoanhydride phosphohydrolase"  
 /protein\_id="AAA0005.1"  
 /db\_xref="GI:145182"  
 /translation="MKAILIPFLSLLIPLPQAFAPQSEBELKESVIVSRHGVAP  
 TKATLMQDVPDAPWTFVKGLWLTFRGGLIAYLGHYQARLVADGLLAKGCPQS  
 GQVAIADVPRTKTEGFAAGLADPCATVHTQADTSPDPLFPLKTVGQCLDNA  
 NVTDAIISRVSGTADTGHRTAFRLERVLNPPQSNCLKREKODESCSLTQALPS  
 ELKVSANSLTGAVSLASMLTEFLQQAQGMPCGGRITDISHQWTLTSLHNAQF  
 YLLORTVEARSRAFLDLIKTALTPHPQKAYGVLTPTSLVFLTAGHDNLNLAGG  
 ALENLWLPQDPNTPGGLVFERWRLSDNQNIQVSLVFQTLQMRDKTFLSLNT  
 PPGEVKTLTAGCEERNAQMCSLAGFTQIVNEARIACSL"  
 299..301  
 /gene="appa"  
 /citation=[1]  
 /phenotype="R16A mutant"  
 variation  
 Query Match 17.0%; Score 225.2; DB 1; Length 1901;  
 Best Local Similarity 53.2%; Pred. No. 7.7e-56;  
 Matches 571; Conservative 0; Mismatches 489; Indels 13; Gaps 4;  
 ORIGIN  
 QY 107 TAGAACGTGTGTTATTTGATCGCATGTGTTCGTCGCGACCAA-ACACGCGC 165  
 Db TGGAAAGTGTGTTATTTGATCGCATGTGTTCGTCGCGACCAA-ACACGCGC 165  
 QY 166 TTATGAATGATTTACGCCAGATAGTGGCCCAATGCGCGTAAAAAGCGGGTATTAA 225  
 Db TGTATGATGATTTACGCCAGATAGTGGCCCAATGCGCGTAAAAAGCGGGTATTAA 225  
 QY 226 CGCCACGTGTGCGGAGTGTGCATGATGGGGGGTATTATGGTGTACTTTCGCA 285  
 Db CACCGCGGGTGTGAGCTATCGCTATCTCGACATTACCAAGCCGCGTGTGTAG 455  
 QY 286 GCCTTGGTTGTAGCG---CGGGGATGTCCGGCAGAGGGGGGGGTATATGCACAGCGAG 342

456 CCGACGATTTGTCGCGAAAAAGGGCTCCCGCAGTCTGGTCAGTCCGATTATTGCTG 515  
 QY 343 ATATCGATCAACGTACTCCGCTTAACCGGACAGGCAATTTCTTGATGGTGTGGTCCGGGGT 402  
 Db 516 ATGTGACGAGGTTACCGGTAACACAGCGGAGCCTTCGCGCGGGGTGGCACTGACT 575  
 QY 403 GTGGTTTACCGGTGATTAATCAGGCGGATTTGAAAAAGACCGATCCCTCTTTCATCCGG 462  
 Db 576 GTGCAATAACCGTACATACCCAGGCGAGATCGTCCAGTCCCGATCCGTTATTTATCTC 635  
 QY 463 TAGAGACTGGCGTGTAAAGTTAGACAACGACAAACAGATAAAGCGATTGAAGAACGAT 522  
 Db 636 TAAAACTGGCGTTTCCGCACTGGATTAACGGAACGACTGACGCGATCTCAGCAGGG 695  
 QY 523 TGGCGCGCGGTAGATACGCTGAGCCAGCGCTACGCTTAAACCTTTTCCGAGATGGGGG 582  
 Db 696 CAGGAGGGTCAATTCGCTACTTTACCGGCGCATCGGCAACCGCGTTTCGCGAATGGAAC 755  
 QY 583 AGTCTGTAATTTTCGGGCTCTCTCTATTCGMAATCTTGCACAGCAAGAAAAAGCT 642  
 Db 756 GGGTGTCTTAATTTTCCGCAATCAAACTGTGCTTAAACGTGAGAAACAGACGAAAGCT 815  
 QY 643 GTGATTTTGGCCACTTTTCGGCCCAATGAAGTTAAGCTTAATAAAGAGGGGCAAAAGTGA 702  
 Db 816 GTTCATTAAGCGCAGCATTTACCATCGAACTCAAGGTGAGCGCGA-----CAATGTCT 869  
 QY 703 CCTCAGTGGGCCACTGGGGTTATCATCGACATTTGGGTGMAATCTTCTTATACAAAG 762  
 Db 870 CATTAAACGGTGGTAAAGCTTCGCTCAATGCTGACGAGATATTTCTCTGCAACAG 929  
 QY 763 CACAAGCCATGCGCAGAGGTTCCTGCGCAACCGCTAAAAAGGGCGGAGAAATTTGGGTATCCT 822  
 Db 930 CACAGGATTCGCGGAGCGGGTGGGAAGATCACCGATTTCACACAGTGAACACCT 989  
 QY 823 TATTGTCTTACATAACCGGCAATTTAATTAATGCGAAAAACACCGTATATCGCGGTC 882  
 Db 990 TGTAAAGTTTGCATAACCGCAATTTTATTTGCTCAACGACGCGCAGGTTGCGCGCA 1049  
 QY 883 ATAAAGGAGCGGCATTATTACAGCAGATAGATACGGCTTAAACCTTCAACTGATGCTC 942  
 Db 1050 GCGCGCCACCCCGTTATTAGATTGATCAAGACAGCGTTGACGCGCCATC---CACCGC 1106  
 QY 943 AGGGGCAAAAGTACCCATTTTCAGGCCCAAAACCGGGTCTTGTCTCGTGGGCGATGATA 1002  
 Db 1107 AAAAAACAGCGGTATGGTGTGACATTTACCCACTTTCAGTGTCTGTTTATCCCGGACACGATA 1166  
 QY 1003 CCAATATTGCCAATATTCGGGTATGCTGGGAGCGGACTGGCAGCTACCCGAGCAACCTG 1062  
 Db 1167 CTAATCTGCAATCTTCGCGCGCGCCTGAGCTCACTGACGCTTCCCGTCAAGCGG 1226  
 QY 1063 ATAATACTCCACGAGTGGGGATTTGGTTTGTAACTATGCAAAATCCGGATAACCACC 1122  
 Db 1227 ATAACAGCGCGCAGGTGGTGAACCTGTTTGAACGCTGGCGTGGCTAAGCGATAACA 1286  
 QY 1123 ACGCTACGTTGCGGTGAGATGTTCTACCAACGATGATCAGTTGGTGTAT 1175  
 Db 1287 GCCAGTGGATTCAGGTTTGGTGTCTTCCAGACTTTACAGCAGATGCGGTAT 1339  
 RESULT 10  
 ECOAPPA  
 LOCUS  
 DEFINITION Escherichia coli periplasmic phosphoanhydride phosphohydrolase  
 (appa) gene, complete cds.  
 ACCESSION M58708  
 VERSION M58708.1 GI:145283  
 KEYWORDS  
 SOURCE Escherichia coli  
 ORGANISM Escherichia coli  
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 Enterobacteriaceae; Escherichia.  
 REFERENCE 1 (bases 1 to 1901)

AUTHORS Dassa, J., Marck, C. and Boquet, P.L.  
 TITLE The complete nucleotide sequence of the *Escherichia coli* gene *appA* reveals significant homology between pH 2.5 acid phosphatase and glucose-1-phosphatase  
 JOURNAL J. Bacteriol. 172 (9), 5497-5500 (1990)  
 MEDLINE 90386616  
 PUBMED 2169385  
 REFERENCE 2 (sites)  
 AUTHORS Ostani, K., Harms, E.H., Stevis, P.E., Kuciel, R., Zhou, M.M. and Van Etten, R.L.  
 TITLE Overexpression, site-directed mutagenesis, and mechanism of *Escherichia coli* acid phosphatase  
 JOURNAL J. Biol. Chem. 267 (32), 22830-22836 (1992)  
 MEDLINE 93054596  
 PUBMED 1429631  
 REFERENCE 3 (sites)  
 AUTHORS Ostani, K. and Van Etten, R.L.  
 TITLE *App304* of *Escherichia coli* acid phosphatase is involved in leaving group protonation  
 JOURNAL J. Biol. Chem. 268 (28), 20778-20784 (1993)  
 MEDLINE 94012612  
 PUBMED 8407904  
 FEATURES Location/Qualifiers  
 1..1901  
 /organism="Escherichia coli"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:562"  
 3..1522  
 /gene="appA"  
 3..8  
 /gene="appA"  
 16..108  
 /gene="appA"  
 /notes="ORF"  
 /codon\_start=1  
 /transl\_table=11  
 /product="unknown"  
 /protein\_id="AA72085.1"  
 /db\_xref="GI:145284"  
 /translation="MWLLWFGVILLMCSLSLTLVWLDPLRLK"  
 118..149  
 /gene="appA"  
 133..138  
 /gene="appA"  
 156..161  
 /gene="appA"  
 172..177  
 /gene="appA"  
 188..1486  
 /gene="appA"  
 /standard\_name="acid phosphatase"  
 /codon\_start=1  
 /transl\_table=11  
 /product="periplasmic phosphoanhydride phosphohydrolase"  
 /protein\_id="AA72086.1"  
 /db\_xref="GI:145285"  
 /translation="TKAIIIFPLSLIPLTPQSAFAQSEPELKLESVIVSRHGVRA  
 PKATQMDVTPAWPTWPKVGLWTPRGELIAYLGHYQRLVADGLAKGCPQS  
 GOVAIADVRGKTEAPAGLA PDCAITVHTQADTSPPDLPNPLTKGCOLDNA  
 NVTDAILSRAGGSIA DFTGHRQTAFLERVLNFPQSNICLKEKODESCSLTQALPS  
 ELKVSADNLSLTGAVSLASMLTEIFLLQQAQMPFPGWGRITDSDHWNTLLSLNAQF  
 YLQRTPEVARSATLLELDLITKALTPHPFKQAYGVITFTSVLFIAGHDTMLNLGG  
 ALBNWTLPGQPTPPGGELVFRWRRLSDNSQWIOVSLVFQTLQOQMDKTPLSLNT  
 PPGEVILTLAQCBERNAAQGMCSLAGFTQIVNEARI PACSL"  
 188..253  
 /gene="appA"  
 254..1483  
 /gene="appA"  
 /product="periplasmic phosphoanhydride phosphohydrolase"  
 299..301  
 /gene="appA"  
 /standard\_name="H303A mutant"  
 /note="created by site directed mutagenesis"

variation /citation=[2]  
 299..301  
 /gene="appA"  
 /standard\_name="R16A mutant"  
 /note="created by site directed mutagenesis"  
 /citation=[2]  
 299..301  
 /gene="appA"  
 /standard\_name="R20A mutant"  
 /note="created by site directed mutagenesis"  
 /citation=[2]  
 299..301  
 /gene="appA"  
 /standard\_name="R92A mutant"  
 /note="created by site directed mutagenesis"  
 /citation=[2]  
 302..304  
 /gene="appA"  
 /standard\_name="H17N mutant"  
 /note="created by site directed mutagenesis"  
 /citation=[2]  
 1163..1165  
 /gene="appA"  
 /standard\_name="D304A mutant"  
 /note="created by site directed mutagenesis"  
 /citation=[3]  
 /phenotype="Asp 304 changed to Ala"  
 /replace="gct"  
 1163..1165  
 /gene="appA"  
 /standard\_name="D304E mutant"  
 /note="created by site directed mutagenesis"  
 /citation=[3]  
 /phenotype="Asp 304 changed to Glu"  
 /replace="gag"  
 1497..1522  
 /gene="appA"  
 /note="rho-independent terminator"  
 terminator  
 ORIGIN  
 Query Match 17.08; Score 225.2; DB 1; Length 1901;  
 Best Local Similarity 53.2%; Pred. No. 7.7e-56;  
 Matches 571; Conservative 0; Mismatches 489; Indels 13; Gaps 4;  
 QY 107 TAGAACGTGTGGTTATTTTGAGTCGCATCGTGTGTCTCGCCGACCAA-ACAACGAGC 165  
 DB 276 TGGAAAGTGTGTGATTGTGTCATGCTGTGTGTGTCTCCAAACCAAGCCACGCAAC 335  
 QY 166 TTATCAATGATGTTACCGCAGATAAGTGCCCAATCGCCGTAAGGCGGGTATTAA 225  
 DB 336 TGATGCAGGATGTCAACCCAGACGATGCGCCCACTGGCCGCTAAACTGGGTGGCTGA 395  
 QY 226 CGCCACGTGTGCGGAGTTGTCACATTGATGGGGGGTTTATGGTGATTACTTTCCGA 285  
 DB 396 CACCGGNGGTGGTGAGCTAATCGCTATCTCGACATTACCAACGCCAGCTCTGGTAG 455  
 QY 286 GCCTTGGTTTGTAGCG---CGCGGATGTCCGCAGAGGGGGGGTATATGCACAGCAG 342  
 DB 456 CCGACGGATTCTCGCGCAAAAAGGGCTGCCCGCAGTGTGTGAGTTCGGTTCGGATTTTGTCTG 515  
 QY 343 ATATCGATCAACGTACCCGCTTAACCGCAGCAGGCAATTTCTTGATGGTGTGGCTCCGGGGT 402  
 DB 516 ATGTGCAGCGGTACCCGTAAACACAGCGGAAGCTTCGCGCGCGGGTGGCACCCTGACT 575  
 QY 403 GTGGTTTACCGTCATTAATCAGGCCGATTTGAAAAGACGATCCCTGTTCATCCGG 462  
 DB 576 GTGCAATAACCGTACATACCCAGGCAGATACGTCCCGATCCGTTATTTAATCTC 635  
 QY 463 TAGAGACTGGCGTGTGTAAGTTAGACACGACACAAAGATAAAGCCGATTAAGAACGAT 522  
 DB 636 TAAAACTGGCGTTTGGCCACTGGATAACGCGAAGTACTGACGCGATCTTCAGCAGGG 695  
 QY 523 TGGCGCGCGCTTAGTACGGTGAGCCAGCGCTACGCTAAACCTTTTGGCCAGATGGGG 582

Db 696 CAGAGGGTCAATTGCTGACTTTACCGGGCATCGGCAACCGGCTTTCGGGAAGTGAAC 755  
Qy 583 ACGTGCTGAATTTTCGGGCTTCTCTTATTGCAAAATCTTTTGCAACGAGGAAACGCT 642  
Db 756 GGGTGTCTTATTTTCGCAATCAAACTTGTCCTTAAACGCTGAGAAACAGGAGAAAGCT 815  
Qy 643 GTGATTTTGGCCACTTTTCGGGCAATCAAGCTTAACGTTAATAAGGAGGACAAAAGTGA 702  
Db 816 GTTCATTAAACGAGGCAATACCAATCGGAATCAAGGTGAGCGCGA-----CAATGTCT 869  
Qy 703 CCCTCAGTGGGCGCACTGGGCTTATCATCGCAATTTGGGTGAATCTCTTATTACAAAACG 762  
Db 870 CATTAACCGGTGCGTGAAGCTCGCATCAATGCTGACGGAGATTTCTCTCGCAACAG 929  
Qy 763 CACAGCCATGCGAGAGTTGCTCGCAACGCTTAAGGGGCGGAGATTTGGGTATCCT 822  
Db 930 CACAGGGAATGCGCGGAGGCTTATGATTTGATCAAGGAGGATCACCGATCACACGAGAACCT 989  
Qy 823 TATTGTCAATACATAACGCGCAATTTAAATTTAATGCGCAAAAACACCGTATATCGCCGCTC 882  
Db 990 TGCTAAGTTGCAATAACGCGCAATTTTATTTGCTACAAACGCAACGAGAGTTGCGCGCA 1049  
Qy 883 ATAAAGGAGCGCCATTTATCAGCAGATAGATAGCGCTTTAAACCTTTCAACTGATGCTC 942  
Db 1050 GCCCGCGCACCCCGCTTATGATTTGATCAAGACAGCGTTGACGCCCATC---CACCGC 1106  
Qy 943 AGGCGCAAAAGCTACCCATTTACGCCCAAAACCGGCTTGTCTCGGTGGGCAATGATA 1002  
Db 1107 AAAAAGCGGTATGCTGATCATCCAGCAACGCAACGCAACGAGAGTTGCGCGCA 1166  
Qy 1003 CCAATATTGCCAATATTGCGGGTATGCTGGAGCGGAGCTGCGAGTACCGAGCAACCTG 1062  
Db 1167 CTAATCTGGCAATCTCGGCGGCACTGGAGCTCAACTGAGACCTTCCCGGTCAGCCGG 1226  
Qy 1063 ATATATCTCCACGAGTGGGGATTTGTTTGAACCTATGCGAAAAATCCGGATAACCAAC 1122  
Db 1227 ATACACGCGCCAGGTGGTGAATGTTGTTTGAACCTGCGTGGCTGAGCGATGAACA 1286  
Qy 1123 AGCGCTAGCTGCGGTGAAGATGTTCTACAAACGATGGATCAGTTCGGTAAT 1175  
Db 1287 GCCAGTGGATTGAGTTTCGCTGCTCTCCAGACTTTACAGCAGATGCGTGAT 1339

RESULT 11  
AX356570 1901 bp DNA linear PAT 06-FEB-2002  
LOCUS AX356570  
DEFINITION Sequence 5 from Patent WO0190333.  
ACCESSION AX356570  
VERSION AX356570.1 GI:18621057  
KEYWORDS Escherichia coli  
SOURCE Escherichia coli  
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.  
REFERENCE 1 Short,J.M., Kretz,K.A., Gray,K.A., Barton,N.R., Garrett,J.B. and O'Donoghue,E.  
AUTHORS Recombinant bacterial phytases and uses thereof  
TITLE Patent: WO 0190333-A 5 29-NOV-2001;  
JOURNAL DIVERSA CORPORATION (US)  
FEATURES Location/Qualifiers  
source 1..1901  
/organism="Escherichia coli"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:562"  
ORIGIN  
Query Match 17.0%; Score 225.2; DB 6; Length 1901;  
Best Local Similarity 53.2%; Pred. No. 7.7e-56;  
Matches 571; Conservative 0; Mismatches 489; Indels 13; Gaps 4;  
Qy 107 TAGAAGCTGTGGTATTATTGAGTCGCCATGGTGTGCTGCGCGCAACAA-ACAGCGCAGC 165

Db 276 TGGAAAGTGTGGTGAATTTGCTAGTGTCTATGTTGGTGTCTTCAACCCAGGCCAGCAAC 335  
Qy 166 TTATGAATGATGTATACGCGCAGATAGTGGCCCAATATGCGCGGTAAAGCGGGGTATTATA 225  
Db 336 TGATGCAAGATGTACACCCAGCAGCATGGCCAACTTGGCCGGTAAACCTGGGTGTGACTGA 395  
Qy 226 CGCCACGCTGGTGGGAGTTGGTCAATTGATGGGGGGTTTTATGGTGAATTTTCGCA 285  
Db 396 CACCCGCGGTGGTGGAGCTAATTCGCTATCTCGACATACCAACGCCAGCGTCTGGTAG 455  
Qy 286 GCCTTGGTTTGTAGCG---CGCGGATCTCGGCGAGAGGGGGGGGTATATGCAAGGAG 342  
Db 456 CCGACGATTTGTCGCGAAAAAGGCTGCCCGCAGTCTGGTCAAGTCGCGATTAATTGCTG 515  
Qy 343 ATATCGATCAACGTTACCGCTTTAACCGGACAGGCAATTTCTTGTATGTTGGTCTCGGGGT 402  
Db 516 ATGTGCAAGAGGCTTACCGGTAAACAGGCGAGCTTTCGCCCGGGGTGCGACCTGACT 575  
Qy 403 GTGGTTTTCACCGTGCATAATCAGGCGGATTTGAAAAAGACCGATCCCTGTTCCATCGG 462  
Db 576 GTGCAATAACCGTACATACCCAGGCGAGATAGTCCAGTCCCGATCCGTTATTTAATCCTC 635  
Qy 463 TAGAGACTGGCGGTGTAGTTAGTAGACAGCAGCAACAGATAAAGCGGATTTGAAGAACAT 522  
Db 636 TAAAAAATGGCGGTTTGGCAACTGGGATAACGGGAACGTGACGTACCGGATCTTCAGCAGG 695  
Qy 523 TGGCGGGGCGGTAGATACGGTGAAGCGCTAGCTTAAACCTTTTCCCGAGATGGGG 582  
Db 696 CAGGAGGGTCAATTGCTGACTTTTACCGGCAATCGGCAAAACGGCGTTTCGCGAAGTGAAC 755  
Qy 583 AGTGTGTGAATTTTCGGGCTTCTCTTATTGCAAACTCTTTCACAGCAGCAAGGAAAAAGCT 642  
Db 756 GGGTGTCTTAATTTTCGCGCAATCAAACTTTGTGCTTTAAACGTGAGAAACAGGACGAAGCT 815  
Qy 643 GTGATTTTGGCCACTTTGGCGCAATGAAGTTAAAGTTAAATAAGAAAGGGAACAAAAGTGA 702  
Db 816 GTTCATTAAACGAGGCAATTTACCTCGGAATCAAGTGAAGCGCGA-----CAATGTCT 869  
Qy 703 CCCTCAGTGGGCACTGGGCTTATCATCGACATTTGGGTGAATCTCTTATTACAAAACG 762  
Db 870 CATTAACCGGTGCGGTAAAGCTCGATCAATGCTCAAGGAGATATTTCTCTGCGAACAG 929  
Qy 763 CACAGCCATGCGAGGTTGCTTGGCAACGCTTAAAGGGGCGGAGAAATTCGGTATCCT 822  
Db 930 CACAGGGAATGCGGAGCGGGGTGGGAGAGATCACCGATTCAACCGATGGAACACCT 989  
Qy 823 TATTGTCAATTAATAACGCGCAATTTAATTTAATGGCAAAAACACGTTATTCGCCGCTC 882  
Db 990 TGCTAAGTTTGCATAACGCGCAATTTTATTTGCTACAAACGCAACGCGAGGTTGCGCGCA 1049  
Qy 883 ATAAAGGAGCGCATTTATACAGCAGATAGATACGCTTTTAACTTCACTGATGCTC 942  
Db 1050 GCGCGCCACCGCTTATTAGATTTGATCAAGACAGGTTGAGCCCATC---CACCGC 1106  
Qy 943 AGGCGCAAAAGCTACCCATTTTCAGCCCAAAACCGGCTTGTCTCTCGTGGGCAATGATA 1002  
Db 1107 AAAAAGCGGCTATGGTGTGACATTACCCACTTCAGTGTCTTATTCGCGGACAGATA 1166  
Qy 1003 CCAATATTGCCAATATTTCGGGTATGCTGGAGCGGAGCTGCGAGTACCGAGCAACCTG 1062  
Db 1167 CTAATCTGGCAATCTCGGCGGCACTGGAGCTCAACTGAGACCTTCCCGGTCAGCCGG 1226  
Qy 1063 ATAACTACTCCACGAGTGGGGATTTGTTTGAACCTATGCGAAAAATCCGGATAACCAAC 1122  
Db 1227 ATAAACGCGCGCAGGTGGTGAATGTTTGAACCTGCGTGGCTGAGCGATGAACA 1286  
Qy 1123 AGCGCTAGCTGCGGTGAAGATGTTCTACAAACGATGGATCAGTTCGGTAAT 1175  
Db 1287 GCCAGTGGATTGAGTTTCGCTGCTCTCCAGACTTTACAGCAGATGCGTGAT 1339

AX356571  
LOCUS AX356571 1901 bp DNA linear PAT 06-FEB-2002  
DEFINITION Sequence 6 from Patent WO0190333.  
ACCESSION AX356571  
VERSION AX356571.1 GI:18621058  
KEYWORDS  
SOURCE Escherichia coli  
ORGANISM Escherichia coli  
REFERENCE  
AUTHORS Short, J.M., Kretz, K.A., Gray, K.A., Barton, N.R., Garrett, J.B. and O'Donoghue, E.  
TITLE Recombinant bacterial phytases and uses thereof  
JOURNAL Patent: WO 0190333-A 6 29-NOV-2001,  
DIVERSA CORPORATION (US)  
FEATURES  
source Location/Qualifiers  
1 .1901  
/organism="Escherichia coli"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:562"

ORIGIN

Query Match 17.0%; Score 225.2; DB 6; Length 1901;  
Best Local Similarity 53.2%; Pred. No. 7.7e-56;  
Matches 571; Conservative 0; Mismatches 489; Indels 13; Gaps 4;

QY 107 TAGAAGCTGTGGTATTATTTAGTGGCGATGGTTCGCTGGCGACCAA-ACAACGCGAGC 165  
DB 276 TGAAGAAGTGGTGGTATGTCAGTCGTCATGGTGGCTCAACCAAGGCGACGCAAC 335  
QY 166 TTATGAATGATGTACGCCAGATAAGTGGCCACAATGGCCGGTAAAGCGGGGTATTAA 225  
DB 336 TGATGAGGATGTCAACCCAGACGATGGCCAACTGGCCGGTAAACCTGGGTGGCTGA 395  
QY 226 CGCACGTGGTGGCGAGTGGTGCATTTGATGGGGGGTATTATGTGTGATTACTTTCGCA 285  
DB 396 CACCGGNGTGGTGGAGTAACTGCTATCTCGACATTACCTGGGCCAGGCTCTGGTAG 455  
QY 286 GCCTTGGTTGTTAGCG---GCGGGATGTCCGGCAGAGGGGGGGGTATATGCAAGCGAG 342  
DB 456 CCGACGGATGTGGCGGAAAGAGGCTGCCCGAGTGTGTCAGTGGTGGGATATTATTCGT 515  
QY 343 ATATCGATCAACGTACCGCTTAACCGGACAGCATTTCTTGATGTGTGGCTCCGGGT 402  
DB 516 ATGTGACAGCGGTACCGGTAACAGGCGAGCCCTTCGCGCGGGCTGGCACCTGACT 575  
QY 403 GTGGTTGACCGTGATTAATCAGGCGGATTTGAAAGAGACCGATCCCTGTTCCATCCGG 462  
DB 576 GTGCATTAACCGTACATACCCAGCGAGATAGTCCAGTCCGATCCGTTATTATTCCTC 635  
QY 463 TAGAGACTGCGGTGTGAAGTTAGACAAACGCAACAGATTAAGCGATGAAGACGAT 522  
DB 636 TAAAAACTGCGTTTGCCTCACTGGATAACCGGAACTGATGACGCGATCCTCAGCAGG 695  
QY 523 TGGCGCGGCGGTAGATACGGTGAGCGCGCTACGCTAAACCTTTTGCCAGATGGGG 582  
DB 696 CAGGAGGTCAATTGCTGACTTTACCGGCGATCGGCAACCGGGTTTCGGGACTGGAAC 755  
QY 583 ACGTGCTGAATTTGGCGCTTCTCTTATTGAAATCTTTGCAACAGCAGGAAAGT 642  
DB 756 GGGTGTCTTAATTTTCGCAATCAAACTTGTGCTTAAACGTGAGAAACAGGACGAAAGT 815  
QY 643 GTGATTTTGCCCACTTTGGCGCAATGAAGTTAAGTAAAGAGGAGCAAAAGTGA 702  
DB 816 GTTCATTAACGCGAGGATTAACATCGGAATCAAGGTGAGCGCGA-----CAATGTCT 869  
QY 703 CCTCAGTGGGCCACTGGCGTTATCATCGACATTTGGGTGAATCTTCTTATTACAAACG 762  
DB 870 CATTAAACCGGTGGGTGAAGCGCTCGCATCAATGTCTGACGGAGATATTCTCTCTGCAAC 929  
QY 763 CACAAGCATGCGCAGAGTTGCTTGGCAACGGCTTAAAGGGGGGAGAAATTTGGGTATCT 822

DB 930 CACAGGAATGCGCGAGCCCGGGTGGGAAGAGATCACCGATTACACCACTGGAAACACT 989  
QY 823 TATTGTCAATTACATAACGCGCAATTTAATTTAATGCAAAACACCGTATATCGCCGTC 882  
DB 990 TGCTAAGTTTGATACGCGCAATTTTATTGCTTACAACGCAACGAGAGGTTGCCGCA 1049  
QY 883 ATAAAGGAGCGCAATTTATTACAGCAGATAGTACGGCTTTAAACCTTCACTGGATGCTC 942  
DB 1050 GCGCGCCACCCCGTTATTAGATTGATCAAGACAGGTTGAGCGCCCATC---CACCG 1106  
QY 943 AGGGGCAAAAGCTACCCATTTTCAGCCCAAAACCGGGTCTTTGTTCTCGTGGCATGATA 1002  
DB 1107 AAAAAGAGCGTATGGTGTGACATTACCACTTCAGTGTGTTTATCGCGGACGATA 1166  
QY 1003 CCATATTGCAATATTGCGGGTATGCTGGAGCGGACGCTGCGAGCTACCGAGCAACCTG 1062  
DB 1167 CTAATCTGGCAAAATCTCGCGGCGCACTGGAGCTCACTGGAGCTTCCCGGTGAGCGG 1226  
QY 1063 ATAATACTCCACAGGTGGGGGATTTGTTTGTGAACTATGCGCAAAATCCGGATAACCA 1122  
DB 1227 ATAACAGCGCGCAGGTGGTGAAGTCTTACCAACGATGATCAGTTCGGTAAT 1286  
QY 1123 AGCCTAGTTCGGTGAAGATGTTCTACCAACGATGATCAGTTCGGTAAT 1175  
DB 1287 GCCAGTGGATTTCAGGTTTCGCTGCTTCCAGACTTTTACAGCAGATGCGTAT 1339

RESULT 13  
AX356572 1901 bp DNA linear PAT 06-FEB-2002  
LOCUS AX356572 Sequence 7 from Patent WO0190333.  
DEFINITION AX356572  
ACCESSION AX356572  
VERSION AX356572.1 GI:18621059  
KEYWORDS  
SOURCE Escherichia coli  
ORGANISM Escherichia coli  
REFERENCE  
AUTHORS Short, J.M., Kretz, K.A., Gray, K.A., Barton, N.R., Garrett, J.B. and O'Donoghue, E.  
TITLE Recombinant bacterial phytases and uses thereof  
JOURNAL Patent: WO 0190333-A 7 29-NOV-2001,  
DIVERSA CORPORATION (US)  
FEATURES  
source Location/Qualifiers  
1 .1901  
/organism="Escherichia coli"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:562"

ORIGIN

Query Match 17.0%; Score 225.2; DB 6; Length 1901;  
Best Local Similarity 53.2%; Pred. No. 7.7e-56;  
Matches 571; Conservative 0; Mismatches 489; Indels 13; Gaps 4;

QY 107 TAGAAGCTGTGGTATTATTTAGTGGCGATGGTTCGCTGGCGACCAA-ACAACGCGAGC 165  
DB 276 TGAAGAAGTGGTGGTATGTCAGTGTGCTCATGGTGGCTCCAAACCAAGGCGACGCAAC 335  
QY 166 TTATGAATGATGTACGCCAGATAAGTGGCCACAATGGCCGGTAAAGCGGGGTATTAA 225  
DB 336 TGATGAGGATGTCAACCCAGACGATGGCCAACTGGCCGGTAAACCTGGGTGGGTGA 395  
QY 226 CGCACGTGGTGGCGAGTGGTGCATTTGATGGGGGGTATTATGTGTGATTACTTTCGCA 285  
DB 396 CACCGGNGTGGTGGAGTAACTGCTATCTCGACATTACCAACGCGAGCGGTCTGGTAG 455  
QY 286 GCCTTGGTTGTTAGCG---GCGGGATGTCCGGCAGAGGGGGGGGTATATGCAAGCGAG 342  
DB 456 CCGACGGATGTGGCGGAAAGAGGCTGCCCGAGTGTGTCAGTGGTGGGATATTATTCGT 515  
QY 343 ATATCGATCAACGTACCGCTTAACCGGACAGCATTTCTTGATGTGTGGCTCCGGGT 402  
DB 516 ATGTGACAGCGGTACCGGTAACAGGCGAGCCCTTCGCGCGGGCTGGCACCTGACT 575  
QY 403 GTGGTTGACCGTGATTAATCAGGCGGATTTGAAAGAGACCGATCCCTGTTCCATCCGG 462  
DB 576 GTGCATTAACCGTACATACCCAGCGAGATAGTCCAGTCCGATCCGTTATTATTCCTC 635  
QY 463 TAGAGACTGCGGTGTGAAGTTAGACAAACGCAACAGATTAAGCGATGAAGACGAT 522  
DB 636 TAAAAACTGCGTTTGCCTCACTGGATAACCGGAACTGATGACGCGATCCTCAGCAGG 695  
QY 523 TGGCGCGGCGGTAGATACGGTGAGCGCGCTACGCTAAACCTTTTGCCAGATGGGG 582  
DB 696 CAGGAGGTCAATTGCTGACTTTACCGGCGATCGGCAACCGGGTTTCGGGACTGGAAC 755  
QY 583 ACGTGCTGAATTTGGCGCTTCTCTTATTGAAATCTTTGCAACAGCAGGAAAGT 642  
DB 756 GGGTGTCTTAATTTTCGCAATCAAACTTGTGCTTAAACGTGAGAAACAGGACGAAAGT 815  
QY 643 GTGATTTTGCCCACTTTGGCGCAATGAAGTTAAGTAAAGAGGAGCAAAAGTGA 702  
DB 816 GTTCATTAACGCGAGGATTAACATCGGAATCAAGGTGAGCGCGA-----CAATGTCT 869  
QY 703 CCTCAGTGGGCCACTGGCGTTATCATCGACATTTGGGTGAATCTTCTTATTACAAACG 762  
DB 870 CATTAAACCGGTGGGTGAAGCGCTCGCATCAATGTCTGACGGAGATATTCTCTCTGCAAC 929  
QY 763 CACAAGCATGCGCAGAGTTGCTTGGCAACGGCTTAAAGGGGGGAGAAATTTGGGTATCT 822

Db 516 ATGTGCGACGAGGTACCGGTAAACAGCGGAGCGCTTGC CGCGCGGCTGGCACCTGACT 575  
Qy 403 GTGTTTACCGTGCATATCAGGCCGATTTGAAAGAGCGATCCCTCTGTTCCATCCGG 462  
Db 576 GTGCAATAACCGTACATACCCAGGCGAGTACGTCCAGTCCCGATCCGTTATTATTCCTC 635  
Qy 463 TAGAGACTGGGTGTGTAAGTACACACGACGACAAACAGATAAGCAATGAAGACGAT 522  
Db 636 TAAAGACTGGGTGTGCAACTGATACCGGACGAGTACGACGCGATCTCCAGCAGGG 695  
Qy 523 TGGGCGGCGCTTAGATACGCTGAGCCAGCGCTACGCTAAACCTTTTGC CAGATGGGG 582  
Db 696 CAGGAGGCTCAATTGCTGACTTTACCGGCGCATCGGCAACGGCGTTTCGCGAATCGAAC 755  
Qy 583 ACGTGCTCAATTTTGGCGCTTCTCTTATTGCAAACTTTTGCAACAGACAGGAAACGTT 642  
Db 756 GGGTGTCTTAATTTTCGCGCAATCAAACTTGTGCTTAAACGCTGAGAAACAGACGAAAGCT 815  
Qy 643 GTGATTTTGGCCACTTTTGC GCGCAATGAAGTTAACTGTTAATAAGAGGAGGACAAAGTGA 702  
Db 816 GTTCATTAACGCGAGCATATTACATCGGACCTCAAGGTGAGCGCGA-----CAATGCTC 869  
Qy 703 CCCTCAGTGGCGCACTGCGGTATATCAGCAATTTGGGTGAAATCTTTTATTACAAAACG 762  
Db 870 CATTAACCGGTGCGGTAAAGCTCGCATCAATGCTGACGAGATATTTCTCTCTGCAACAG 929  
Qy 763 CACAAGCATGCGCAGAGGTTGCTTGGCAACCGCTAAAGGGGCGGAGAAATTTGGTATCCT 822  
Db 930 CACAGGNAATGCGGAGCGGGTGGGAGGATCACGATTCACACAGTGGAACACT 989  
Qy 823 TATTGTCAATACACGCGCAATTTAATTTAATGGAACAAACCGCTATATCGCCGTC 882  
Db 990 TGCTAAGTTTGCATAACGCGCAATTTTATTGCTTACAAACGCAACCCAGAGGTTGCCGCA 1049  
Qy 883 ATAAAGGAGCGCCATTATCAGCAGATAGTACGGCTTTAAACCTTCAACTGGATGCTC 942  
Db 1050 GCGCGGCAACCGCTTATTAGATTGATCAAGCAGCGTTGACGCCCATC---CACCGC 1106  
Qy 943 AGGGGCAAAAGCTACCATTTTCAGCCCAAAACCGGCTCTTGTCTCGTGGGCGATGATA 1002  
Db 1107 AAAAAGCGGCTATGGTGTGACATTAACCACTTCACTGCTGTTTATCGCGGACACGATA 1166  
Qy 1003 CCATATGCAATATTGCGGTATGCTGGAGCGGAGTGGAGCTACCCGAGCAACCTG 1062  
Db 1167 CTAATCTGGCAAAATCTCGGCGGCGCACTGGAGCTCAACTGGAGCTTCCCGGTGAGCCGG 1226  
Qy 1063 ATAATACTCCACAGGTGGGGATGGTTTGTGAACTATGCGCAAAATCCCGGATAACCAAC 1122  
Db 1227 ATAACAGCGCGCAGGTGGTGAACCTGGTGTGAACTGCGCTGCGCTAAGCGATAACA 1286  
Qy 1123 AGCCTAGTTCGCGTGAAGATGTTCTACCAAGCATGGATCAGTTGCGTAAT 1175  
Db 1287 GCCAGTGGATTGAGTTTGCCTGCTCTCCAGACTTTTACAGCAGATGCGTAT 1339

RESULT 14  
AX042376  
LOCUS AX042376 3470 bp DNA linear PAT 23-NOV-2000  
DEFINITION Sequence 5 from Patent WO0064247.  
ACCESSION AX042376  
VERSION AX042376.1 GI:11340994  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM synthetic construct  
1 artificial sequences.  
REFERENCE  
1 Forsberg,C.W., Golovan,S. and Phillips,J.P.  
AUTHORS Transgenic animals expressing salivary proteins  
TITLE Patent: WO 0064247-A 5 02-NOV-2000;  
JOURNAL UNIVERSITY OF GUELPH (CA)  
FEATURES  
1..3470 Location/Qualifiers  
source /organism="synthetic construct"

/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="R15/APPA transgene"

Query Match 17.0%; Score 224.6; DB 6; Length 3470;  
Best Local Similarity 53.2%; Pred. No. 1.2e-55;  
Matches 571; Conservative 0; Mismatches 489; Indels 13; Gaps 4;

ORIGIN

Qy 107 TAGAAGCTGTGTTATTTTGTAGTCCGCATGTTGCTCGCGGACCAA-ACAAGCGACG 165  
Db 1899 TGGAAAGTGTGTTGATTTGTCAGTCTGTTGTTGCTGCTCAACCAAGGCGCACGCAAC 1958  
Qy 166 TTATGATGATGTTTACGCGCAGATAGTGGCCCAATAGCCGCTAAAGCGGGGTATTAA 225  
Db 1959 TGATGAGATGTCAACCCAGACGATGCGCAACCTGCGCGGTAAACCTGGGTGGCTGA 2018  
Qy 226 CCGCAGCTGCTGCGAGTTGGTCAATGATCGGGGGTTTTATGTTGATTACTTTTCGCA 285  
Db 2019 CACCGCGGTGTTGAGCTAATCGCTATCTCGGACATTTACCAACGCGCTCTGGTAG 2078  
Qy 286 GCCTTGTGTTTGTAGCG---GCGGATGTCCGCGAGAGGGGGGGGTATATGACAGGCGAG 342  
Db 2079 CCGACGAGTTGTGCGGAAAGGGCTGCCCGCAGTCTGCTCAGGTCCGCAATTTTGTGTG 2138  
Qy 343 ATATCGATCAACGATACCGCTTTAAACGCGACAGGCAATTTCTTGATGTTGCTCGGGGT 402  
Db 2139 ATGTGCAAGAGGTACCGGTAAACAGGGGAGCCCTTGC CGCGCGGCTTGGCACCTGACT 2198  
Qy 403 GTGTTTGTGCTGTGATATATCAGGCGGATTTGAAAGAGACCGATCCCTGTTCCATCCGG 462  
Db 2199 GTGCAATAACCTGATACATACCCAGGCGAGATACGTCAGTCCCGATCCGTTATTATCCTC 2258  
Qy 463 TAGAGACTGGCGTGTGTAAGTTAGACAGCAGCAACAGATAAACCGATTGAAGACCAT 522  
Db 2259 TAAAGACTGGCTTTGCCAACTGGATTAACGCAAGCTGACGCGATCTCCACAGGG 2318  
Qy 523 TGGCGGGCGGTAGATACGCTGAGCGCTAGCTAAACCTTTTCCAGATGGGG 582  
Db 2319 CAGGAGGGTCAATGCTGACTTTTACCGGCGCATCGGCAACCGCGCTTTCGCAACTGGAAC 2378  
Qy 583 AGTCTGATTTTGGGCTTCTCTCTTATTCGAATCTTTGCAACAGCAAGGAAACGT 642  
Db 2379 GGGTCTTAATTTTCCGCAATCAAACTTGTGCTTTAAACGTTGAGAAACAGGACGAAAGCT 2438  
Qy 643 GTGATTTTGCCTTTCGCGCAATGAAGTTAACTGTTAATAAGAGGAGCAAAAGTGA 702  
Db 2439 GTTCATTAACGCGAGCATTTACATCGGAACCTCAAGGTGAGCGCGA-----CAATGCTC 2492  
Qy 703 CCTCAGTGGGCGCTGCGGTATTCATCGACATTTGGGTGAAATCTTCTTATTACAAACG 762  
Db 2493 CATTAACCGGTGCGGTAAAGCTTCGATCAATGCTGACGAGATATTTCTCTGCAACAG 2552  
Qy 763 CACAAGCCATGCGCAGAGTTGCTCGCAACCGCTAAAGGGGCGGAGAAATTTGGGTATCCT 822  
Db 2553 CACAGGATGCGGAGCGGGTGGGAGGATCACCGATTTCACACGATGGACACT 2612  
Qy 823 TATTGTCAATTAATAAGCGCAATTTAATTTAATGGCAAAACACCGTATATTCGCCCGTC 882  
Db 2613 TGTAAAGTTTGCATAACCGCAATTTTATTTGCTACAAACGCGCAGAGGTTGCCGCA 2672  
Qy 883 ATAAAGGAGCGCATTTATTACAGCAGATAGATAGCGCTTTAACCTTCAACTGGATGCTC 942  
Db 2673 GCGGCGCAACCGCTTATTAGTTTGTATCAAGACAGCGTTGACGCCCATC---CACGCG 2729  
Qy 943 AGGGGCAAAAGCTACCCATTTTCAGCCCAAAACCGGCTCTTGTCTCGGTGGGCTATGATA 1002  
Db 2730 AAAAAGCGGCGTATGCTGTGACATTTACCCACTTTCAGTGTCTGTTTATCGCGGACACGATA 2789  
Qy 1003 CCAATATTGCCAATATTGCGGTATGCTGGGAGCGGATGCGCAGCTACCCGAGCAACCTG 1062  
Db 2790 CTAACTGCGCAAACTCTCGCGCGCGCACTGGAGCTCAACTGGAGCTTCCCGGTACGCCGG 2849



Thu May 6 08:41:39 2004

QY	1063	ATAAATCTCCACGAGTGGGGANTGGTTTTGAACATATGGCAAAATCGGATAACCACC	1122
Db	2850	ATAAACACGCCGCAGGTGGTGAACTGGTGTGTTTGAACGCTGGCGTCGCTTAAGCGATAACA	2909
QY	1123	AGCGCTACGTTCGGGTGAAGATCTTCTACCAACGATGATCAGTTGGCGTAAT	1175
Db	2910	GCCAGTGGATTCAAGTTTCGCTGGTCTTCGACTTTACAGCAGATCGGTAT	2962
RESULT 15			
AX042374			
LOCUS	AX042374	4060 bp	DNA linear PAT 23-NOV-2000
DEFINITION	Sequence 3 from Patent WO0064247.		
ACCESSION	AX042374		
VERSION	AX042374.1	GI:11340992	
KEYWORDS	synthetic construct		
SOURCE	synthetic construct		
ORGANISM	artificial sequences.		
REFERENCE	1.		
AUTHORS	Forsberg,C.W., Golovan,S. and Phillips,J.P.		
TITLE	Transgenic animals expressing salivary proteins		
JOURNAL	Patent: WO 0064247-A 3 02-NOV-2000;		
	UNIVERSITY OF GUELPH (CA)		
FEATURES	Location/Qualifiers		
source	1..4060		
	/organism="synthetic construct"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:32630"		
	/note="R15/APPA + intron transgene"		
ORIGIN			
	Query Match	17.0%; Score 224.6; DB 6; Length 4060;	
	Best Local Similarity	53.2%; Pred. No. 1.2e-55;	
	Matches 571; Conservative	0; Mismatches 489; Indels 13; Gaps 4;	
QY	107	TAGAACGTGTGGTTATTTTGAGTCGCANGGTGTTTCGTCGCCGACCAA-ACAACGCGAC	165
Db	1899	TGGAAAAGTGTGTGATGTTCAGTCGTGTCATGTTGTGGTGTCTCCAAACCAAGGCCACGCAAC	1958
QY	166	TTATGATGATGTTACGCCAGATAAGTGGCCACATATGCGCGTAAAGCGGGGTATTTAA	225
Db	1959	TGATGACGAGATGTACCCCCAGCATGTGGCCAATCTGGCGCGTAAACTGGGTGGCTGA	2018
QY	226	CGCCACGTGTGCGGAGTTGGTCACATTTGATGGGGGGTTTTATGTGTGTTACTTTTCGCA	285
Db	2019	CACCGCGCGTGTGAGCTAATCGCCTATCTCGACATTACCAACGCCGCTCTGTTAG	2078
QY	286	GCCTTGGTTTGTAGCG---GCGGATGTCCGSCAGNGGGGGGTATATGCACAGCAG	342
Db	2079	CCGACGATTGCTGGCGAAAAAGGGCTGCCCGCAGTCTGTGTCAAGTCGCGATTATTGCTG	2138
QY	343	ATATCGATCAACGTATCCCGCTTAACCGGACAGGCATTCTTGTGTTGGCTCCGGGGT	402
Db	2139	ATGTGACGAGCGTACCCGTAACACAGCGAAGCCTTCGCCGCGGGCTGGCACTTGACT	2198
QY	403	GTGCTTTGACCGTGCATATAATCAGCGCCGATTTGAAAAGAACGATFCCCTCTTCCATCCGG	462
Db	2199	GTGCAATAACCGGTACATACCCAGGACATAGCTCCAGTCGCCGATCCGTTATTTAATCCCTC	2258
QY	463	TAGAGACTGCGCTGTGAAGTTAGACAAACGACAAACAGATAAAGCGATTGAAGAACGAT	522
Db	2259	TAAAACTGGCTTTTGCACTGATTAACGGGAACGTGACTGACGCGATCTCAGCAGGG	2318
QY	523	TGGCGCGGCGCTTAGATACGGTGAGCGAGCGCTACCGCTAAACCTTTTGGCCAGATGGGG	582
Db	2319	CAGGAGGCTCAATTGCTGACTTTTACCGGGCATCGGCAACGGCGTTTTCGCGAACTGGAAC	2378
QY	583	ACGTGCTGAATTTTGGCGCTTCCTTTATGCAAACTCTTTGCAACGACGAGGAACACGT	642
Db	2379	GGGTGCTTAATTTTTCGCAATCAAACTTGTGCTTTAAACGCTGAGGAACACGAGT	2438
QY	643	GTGATTTTGGCCACTTTTGGCGCCCAATGAAGTTTACGTTTAAATAAGAGGACAAAAGTGA	702



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2004, 19:58:08 ; Search time 378.238 seconds

(without alignments)  
14881.793 Million cell updates/sec

Title: US-10-021-723B-3

Perfect score: 1325

Sequence: 1 atgtcagattagaaatcg.....aacctgctgccacatctaa 1325

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 337363 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 29Jan04.\*

1: Geneseq1980s.\*

2: Geneseq1990s.\*

3: Geneseq2000s.\*

4: Geneseq2001as.\*

5: Geneseq2001bs.\*

6: Geneseq2002s.\*

7: Geneseq2003as.\*

8: Geneseq2003bs.\*

9: Geneseq2003cs.\*

10: Geneseq2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1314	99.2	1326	6	ABQ73843	Abq73843 Yersinia
2	1306	98.6	1326	7	ACA53738	ACA53738 Prokaryot
3	1304.2	98.4	1326	6	ABQ73842	Abq73842 Yersinia
4	827.6	62.5	1326	6	ABQ73847	Abq73847 Yersinia
5	242	18.3	1281	8	ACF35789	ACF35789 Nov9X phy
6	226.4	17.1	5421	3	AAC68299	AAC68299 SV40/APPA
7	226.2	17.1	1308	9	ADC87742	ADC87742 DNA encod
8	225.2	17.0	1901	6	AD25463	AD25463 Escherich
9	225.2	17.0	1901	7	ADA19452	Ada19452 E. coli K
10	225.2	17.0	1901	7	ADA19449	Ada19449 E. coli K
11	225.2	17.0	1901	9	ADC87744	ADC87744 DNA encod
12	224.6	17.0	1299	3	ABK12514	ABK12514 DNA encod
13	224.6	17.0	1299	7	ACA19297	ACA19297 Prokaryot
14	224.6	17.0	3470	3	AAC68298	AAC68298 R15/APPA
15	224.6	17.0	4060	3	AAC68296	AAC68296 R15/APPA
16	224.6	17.0	6116	3	AAC68297	AAC68297 R15/APPA
17	224.6	17.0	6708	3	AAC68295	AAC68295 R15/APPA
18	224.6	17.0	17732	3	AAC68300	AAC68300 Lama2/APP
19	224.6	17.0	20623	3	AAC68294	AAC68294 Lama2/APP
20	221.4	16.7	1486	4	AAD06831	AAD06831 E. coli a
21	218.8	16.5	1323	5	AAC88985	AAC88985 Escherich
22	218.8	16.5	1323	6	AAD36473	AAD36473 Escherich
23	218.8	16.5	1323	6	AAD25460	AAD25460 Escherich

24	218.8	16.5	1323	7	ADA19445	Ada19445 E. coli B
25	218.2	16.5	1486	4	AAD06832	AAD06832 E. coli a
26	218.2	16.5	1486	7	ACC57673	ACC57673 Escherich
27	218.2	16.5	1489	3	AAA28216	AAA28216 E. coli a
28	218.2	16.5	1489	7	ACC57672	ACC57672 Escherich
29	177.8	13.4	1256	8	AAD57147	AAD57147 Maize-opt
30	177.8	13.4	1313	8	AAD57148	AAD57148 pNOV4054
31	177.4	13.4	1331	8	ABQ73846	Abq73846 Yersinia
32	177.4	13.4	1431	6	ABQ73846	Abq73846 Yersinia
33	143.6	10.8	1272	2	AAX26540	AAX26540 DNA encod
34	133.2	10.1	846	4	AAA41940	AAA41940 Genomic s
35	104.4	7.9	1737	5	AAS89885	AAS89885 DNA encod
36	100.4	7.6	519	4	AAA41139	AAA41139 CDNA encod
37	73.2	5.5	466	7	ACA15311	ACA15311 Prokaryot
38	68.4	5.2	11710	6	AAH43718	AAH43718 E. coli g
39	65.8	5.0	1264	5	AAS88443	AAS88443 DNA encod
40	65.8	5.0	1264	6	AAS94316	AAS94316 DNA encod
41	64.6	4.9	1230	6	ABQ73845	Abq73845 Rhizobium
42	59.8	4.5	1266	6	ABQ73844	Abq73844 Yersinia
43	50.2	3.8	2000	7	ADA71938	Ada71938 Rice gene
44	48.2	3.6	4050	5	AAS64582	AAS64582 DNA encod
45	48	3.6	5975	2	AAQ55758	AAQ55758 Klebsiell

#### ALIGNMENTS

#### RESULT 1

ABQ73843  
ID ABQ73843 standard; DNA; 1326 BP.  
XX  
AC ABQ73843;  
XX  
DT 08-OCT-2002 (first entry)  
XX  
DE Yersinia pestis phytase nucleotide sequence SEQ ID NO:3.  
XX  
KW Yersinia pestis; phytase; enzyme; E.C.3.1.3.8; phytate; foodstuff; feed;  
KW gene; ds.  
XX  
OS Yersinia pestis.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..1260  
FT /tag= a  
FT /partial  
FT /EC\_number= "3.1.3.8"  
FT /product= "phytase"  
FT /note= "no stop codon given"

XX WO200248332-A2.  
XX  
XX  
XX  
XX 20-JUN-2002.  
XX  
XX 12-DEC-2001; 2001WO-US048774.  
XX  
XX 12-DEC-2000; 2000US-0255090P.  
XX (DIVE-) DIVERSA CORP.  
XX  
XX Short J, Mathur EJ, Richardson T, Robertson D, Barton N;  
XX  
XX WPI; 2002-583504/62.  
XX P-PSDB; ABP51930.  
XX  
XX Novel recombinant phytase protein and polynucleotide for improving  
XX nutritional value of phytate-containing foodstuff, in animal feed and  
XX feed supplements and to degrade excess phytase from environment or  
XX sample.  
XX Claim 1; Fig 5C; 208pp; English.  
XX  
XX The present invention describes an isolated phytase protein (I). (I) can

CC be used for improving the nutritional value of a phytate-containing  
CC foodstuff, by contacting the phytate-containing foodstuff with (i), where  
CC the enzyme catalyses the liberation of inorganic phosphate from the  
CC phytate-containing foodstuff, and so improving the nutritive value of the  
CC contacted foodstuff. The liberation of inorganic phosphate occurs prior  
CC to or after the ingestion of phytate-containing foodstuff by a recipient  
CC organism. Nucleotide sequences (ii) encoding (i) can be used for  
CC producing an animal feed, by transforming a plant, plant portion or plant  
CC cell with a nucleic acid expression vector, comprising (ii), culturing  
CC the plant, plant portion or plant cell under conditions in which the  
CC phytase protein is expressed and converting the plant, plant portion or  
CC plant cell into a composition suitable for animal feed. The animal is  
CC preferably a monogastric animal or a ruminant. (i) and (ii) are useful in  
CC animal feed and feed supplements as well as in treatments to degrade or  
CC remove excess phytate from the environment or a sample. (i) reduces  
CC phytate levels in animal manure and so reduces the phosphate pollution of  
CC the environment. The present sequence encodes a phytase from the present  
CC invention  
XX  
SQ Sequence 1326 BP; 351 A; 285 C; 359 G; 331 T; 0 U; 0 Other;

Query Match 99.2%; Score 1314; DB 6; Length 1326;  
Best Local Similarity 99.9%; Pred. NO. 0;  
Matches 1325; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
QY 1 ATGTCAGTATTAGAAATCGTGTACGGCTATCTGGCTGTGTATGATGCTAAGCGGATG 60  
DB 1 ATGTCAGTATTAGAAATCGTGTACGGCTATCTGGCTGTGTATGATGCTAAGCGGATG 60  
QY 61 GCTGCTATTACTCGCGGTAGCCGCGAGCCATCGGCTATCTTTAGAACGTGTGTT 120  
DB 61 GCTGCTATTACTCGCGGTAGCCGCGAGCCATCGGCTATCTTTAGAACGTGTGTT 120  
QY 121 ATTTTGAGTCGCCATGCTGTTCCTCGCGAGCCAAAC-AACGAGCTTATGATGATGTT 179  
DB 121 ATTTTGAGTCGCCATGCTGTTCCTCGCGAGCCAAAC-AACGAGCTTATGATGATGTT 180  
QY 180 ACGCCAGATAAGTGCCGCAATGCGCGGTAAAGCGGGGTATTTAAGCCACGTGTGCG 239  
DB 181 ACGCCAGATAAGTGCCGCAATGCGCGGTAAAGCGGGGTATTTAAGCCACGTGTGCG 240  
QY 240 GAGTTGCTCATTTGATGGGGGGTTTTATGGTGATTCTTTTCGAGCCTTGTTGTTA 299  
DB 241 GAGTTGCTCATTTGATGGGGGGTTTTATGGTGATTCTTTTCGAGCCTTGTTGTTA 300  
QY 300 GCGCGGGGATGTCGGCAGAGGGGGGGGTATATGCAAGCAGATATCGATCAAGTACC 359  
DB 301 GCGCGGGGATGTCGGCAGAGGGGGGGGTATATGCAAGCAGATATCGATCAAGTACC 360  
QY 360 CGCTTAACCGGACAGGCAATTTCTTGATGGTGCGCTCCGGGGTGTGTTGACCGTGCAT 419  
DB 361 CGCTTAACCGGACAGGCAATTTCTTGATGGTGCGCTCCGGGGTGTGTTGACCGTGCAT 420  
QY 420 AATCAGCCGATTGTGAAAAAGACCGATCCCTGTTCCATCCGCTAGAGACTGGCGTGT 479  
DB 421 AATCAGCCGATTGTGAAAAAGACCGATCCCTGTTCCATCCGCTAGAGACTGGCGTGT 480  
QY 480 AAGTTAGACACGACCAACAGATTAAGCGATTGAGAACGATTGGCGGGCCGTTAGAT 539  
DB 481 AAGTTAGACACGACCAACAGATTAAGCGATTGAGAACGATTGGCGGGCCGTTAGAT 540  
QY 540 ACGGTGAGCCAGCGCTACGCTAAACCTTTTCCGAGATGGGGGACGCTGCTGAATTTGCG 599  
DB 541 ACGGTGAGCCAGCGCTACGCTAAACCTTTTCCGAGATGGGGGACGCTGCTGAATTTGCG 600  
QY 600 GCTTCTCTTATTGCAATCTTTGCAACGACGAGGAAACGTTGATTTTGGCCACTTT 659  
DB 601 GCTTCTCTTATTGCAATCTTTGCAACGACGAGGAAACGTTGATTTTGGCCACTTT 660  
QY 660 GCGGCCAATGAAGTTAAGCTTAATAAGAGAGGACAAAGTGAACCTTCAGTGGGCCACTG 719  
DB 661 GCGGCCAATGAAGTTAAGCTTAATAAGAGAGGACAAAGTGAACCTTCAGTGGGCCACTG 720

QY 720 GCGTTATCATCGACATTGGGTGAATCTTCTTATTATCAAAACGCAAGCCATGCCAGAG 779  
DB 721 GCGTTATCATCGACATTGGGTGAATCTTCTTATTATCAAAACGCAAGCCATGCCAGAG 780  
QY 780 GTTCCCTGGCAACGGCTAAAAAGGGGGGAGAAATGGGTATCCTTATTGTTCATTACATAAC 839  
DB 781 GTTCCCTGGCAACGGCTAAAAAGGGGGGAGAAATGGGTATCCTTATTGTTCATTACATAAC 840  
QY 840 GCGCAATTTAATTAATGGCAAAACACCGCTATATGCGCCGTATTAAGAGGCGCCATT 899  
DB 841 GCGCAATTTAATTAATGGCAAAACACCGCTATATGCGCCGTATTAAGAGGCGCCATT 900  
QY 900 TTACAGCAGATAGATACGGCTTTTAAACCCCTTCAACTGTGATCTCAGGGGCAAAAGCTACCC 959  
DB 901 TTACAGCAGATAGATACGGCTTTTAAACCCCTTCAACTGTGATCTCAGGGGCAAAAGCTACCC 960  
QY 960 ATTTACGCCCAAAACCGGGTCTTGTTCCTCGGTGGGATGATACCAATATTTGCCAATATT 1019  
DB 961 ATTTACGCCCAAAACCGGGTCTTGTTCCTCGGTGGGATGATACCAATATTTGCCAATATT 1020  
QY 1020 GCGGGTATGCTGGGAGCCGACTGGCAGCTACCCGAGCAACCTGATATATCTCCACCAGGT 1079  
DB 1021 GCGGGTATGCTGGGAGCCGACTGGCAGCTACCCGAGCAACCTGATATATCTCCACCAGGT 1080  
QY 1080 GGGGGATTGGTTTTTGAACCTATATGGCAAAATCCCGATTAACACCCAGCGCTACGTTGCGGTG 1139  
DB 1081 GGGGGATTGGTTTTTGAACCTATATGGCAAAATCCCGATTAACACCCAGCGCTACGTTGCGGTG 1140  
QY 1140 AAGATGTTCTACCAACGATGGATCAGTTCGTAATGCCGAAATTTGGATCTGAAAAAT 1199  
DB 1141 AAGATGTTCTACCAACGATGGATCAGTTCGTAATGCCGAAATTTGGATCTGAAAAAT 1200  
QY 1200 AACCCAGCGGGTATTATTTCCTGTCAGTTGCTGTTGTGAAAAATACCGTGACGATAAG 1259  
DB 1201 AACCCAGCGGGTATTATTTCCTGTCAGTTGCTGTTGTGAAAAATACCGTGACGATAAG 1260  
QY 1260 CTTTTCGAGCTTGATACATTCCCAAGAGAGTGGCTAAGGTAAATTTGAACCTGCTGCCAC 1319  
DB 1261 CTTTTCGAGCTTGATACATTCCCAAGAGAGTGGCTAAGGTAAATTTGAACCTGCTGCCAC 1320  
QY 1320 ATCTAA 1325  
DB 1321 ATCTAA 1326

RESULT 2  
ACAS3738  
ID ACA53738 standard; DNA; 1326 BP.  
XX ACA53738;  
AC ACA53738;  
DT 19-JUN-2003 (first entry)  
XX Prokaryotic essential gene #35395.  
DE Antisense; ds; prokaryotic essential gene; cell proliferation;  
KW drug design; gene.  
XX Yersinia pestis.  
OS Yersinia pestis.  
XX WO200277183-A2.  
FN 03-OCT-2002.  
XX 21-MAR-2002; 2002WO-US009107.  
XX 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
DR P-PSDB; ASU49868.  
XX  
XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
XX Claim 14; SEQ ID NO 41608; 1766pp; English.

CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: the sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences

Sequence 1326 BP; 348 A; 286 C; 361 G; 331 T; 0 U; 0 Other;

Query Match 98.6%; Score 1306; DB 7; Length 1326;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 1320; Conservative 0; Mismatches 5; Indels 1; Gaps 1;  
1 ATGTCAGTATTAGAAATCGGTACGGCTATCTGGCTGGTATTGATGCTAAGCGGATTG 60  
1 ATGTCAGTATTAGAAATCGGTACGGCTATCTGGCTGGTATTGATGCTAAGCGGATTG 60  
61 GCTGCTATTACTGCGCGGTTAGCGCGGAGCCATCGGGCTATATCTTAGAACGTGGTT 120  
61 GCTGCTATTACTGCGCGGTTAGCGCGGAGCCATCGGGCTATATCTTAGAACGTGGTT 120  
121 ATTTTGGTCCCATGGTGTTCGCTCCCGACCAAC- AACGACGCTTATGATGATT 179  
121 ATTTTGGTCCCATGGTGTTCGCTCCCGACCAACAAACGACGCTTATGATGATT 180  
180 ACGCCAGATAAGTGCCCAATAGCCCGGTAAAGCGGGTATTAAACGCCACGTGGTGG 239  
181 ACGCCAGATAAGTGCCCAATAGCCCGGTAAAGCGGGTATTAAACGCCACGTGGTGG 240  
240 GAGTTGGTCCATTGATGGGGGGTATTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 299  
241 GAGTTGGTCCATTGATGGGGGGTATTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300  
300 CGCGCGGATGCTCGCGCAGAGGGGGGTATATGACAGGAGGATATCGATCAACGTACC 359

Db 301 GCGCGGGATGTCGCGCAGAGGGGGGTATATGACAGGAGATATCGATCAACGTACC 360  
Qy 360 CGCTTAACCGGACAGGCAATCTTCTGATGGTGGCTCGGGGGTGGTGGTGGAT 419  
Db 361 CGCTTAACCGGACAGGCAATCTTCTGATGGTGGCTCGGGGGTGGTGGTGGAT 420  
Qy 420 AATCAGCGCGATTTGAAAAGACCGATCCCTGTTCCATCCGTTAGAGATCGCGGTGT 479  
Db 421 AATCAGCGCGATTTGAAAAGACCGATCCCTGTTCCATCCGTTAGAGATCGCGGTGT 480  
Qy 480 AAGTTAGACACGCAACACAGATTAAGGATTAAGACGATTAAGGATTAAGGATTAAG 539  
Db 481 AAGTTAGACACGCAACACAGATTAAGGATTAAGGATTAAGGATTAAGGATTAAG 540  
Qy 540 ACGGTGAGCGCGCTACGCTAAACCTTTTGGCCAGATGGGGAGCGTGTGTAATTTGG 599  
Db 541 ACGGTGAGCGCGCTACGCTAAACCTTTTGGCCAGATGGGGAGCGTGTGTAATTTGG 600  
Qy 600 GCTTCTCTTATGCAATCTTTGCAACAGCAAGGAGAAACGCTGATTTTCCACATTT 659  
Db 601 GCTTCTCTTATGCAATCTTTGCAACAGCAAGGAGAAACGCTGATTTTCCACATTT 660  
Qy 660 GCGGCCAATGAAGTTAAAGCTTAATAAGGAGGACAAAAGTCACTCACTGGGCGCAT 719  
Db 661 GCGGCCAATGAAGTTAAAGCTTAATAAGGAGGACAAAAGTCACTCACTGGGCGCAT 720  
Qy 720 GGTATATCATCGACATGCGTGAATCTTTTATTAACAAACGCAAGGAGGATGCGAG 779  
Db 721 GGTATATCATCGACATGCGTGAATCTTTTATTAACAAACGCAAGGAGGATGCGAG 780  
Qy 780 GTTGCTGCGCAACGCTAAAGGCGGAGGATTCGGTATCTTATTTGCTATTACATAAC 839  
Db 781 GTTGCTGCGCAACGCTAAAGGCGGAGGATTCGGTATCTTATTTGCTATTACATAAC 840  
Qy 840 GCGCAATTTAATTAATGCAAAAAACCGTATATCGCCCGTCAATAAGGAGGAGGAT 899  
Db 841 GCGCAATTTAATTAATGCAAAAAACCGTATATCGCCCGTCAATAAGGAGGAGGAT 900  
Qy 900 TTACGACAGATAGATACGCTTTAAACCTTCACTGATGCTCAGGGGCAAGGATACCC 959  
Db 901 TTACGACAGATAGATACGCTTTAAACCTTCACTGATGCTCAGGGGCAAGGATACCC 960  
Qy 960 ATTTTCAGCGCCAAACCGGGTCTTTGTTCTCGTGGGCGATGATACCAATATTGCGCAAT 1019  
Db 961 ATTTTCAGCGCCAAACCGGGTCTTTGTTCTCGTGGGCGATGATACCAATATTGCGCAAT 1020  
Qy 1020 GCGGTATGCTGGGAGCGGCTGCGGAGCTACCGGAGCAACCTGATTAATCTCCACAGGT 1079  
Db 1021 GCGGTATGCTGGGAGCGGCTGCGGAGCTACCGGAGCAACCTGATTAATCTCCACAGGT 1080  
Qy 1080 GCGGGATTTGGTTTTTGAACCTATGCGCAAAATCCGGATAACCAACGCGCTACGTTGCGGTG 1139  
Db 1081 GCGGGATTTGGTTTTTGAACCTATGCGCAAAATCCGGATAACCAACGCGCTACGTTGCGGTG 1140  
Qy 1140 AAGATGTTTACCAACGATGATCAGTTGGTATGCGGAAAAATTTGATCTGAAAAAT 1199  
Db 1141 AAGATGTTTACCAACGATGATCAGTTGGTATGCGGAAAAATTTGATCTGAAAAAT 1200  
Qy 1200 AACCAGCGGGTATTATTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1259  
Db 1201 AACCAGCGGGTATTATTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260  
Qy 1260 CTTTGGAGCTTGATACATTTCAAAAGAAAGTGGCTAAGTAAATTTGAACCTGCTGCCAC 1319  
Db 1261 CTTTGGAGCTTGATACATTTCAAAAGAAAGTGGCTAAGTAAATTTGAACCTGCTGCCAC 1320  
Qy 1320 ATCTAA 1325  
Db 1321 ATCTAA 1326

RESULT 3  
AB073842

1 ATGTCAGTATTAGAAAATCGTGACGGCTATCTGGCGTGGTATTGATGCTAAAGCGGATG 60  
61 GCTGCTATTACTGCGCGGCTAGCCGCCAGCCATCGGCTATATCTTTAGAACGTTGGGTT 120  
61 GCTGCTATTACTGCGCGGCTAGCCGCCAGCCATCGGCTATATCTTTAGAACGTTGGGTT 120  
121 ATTTTGGAGTCCCATGGTGTGGCTCGCCGACCAAC-AAAGCGAGCTTATGAATGATGTT 179  
121 ATTTTGGAGTCCCATGGTGTGGCTCGCCGACCAACAAACGAGCTTATGAATGATGTT 180  
180 ACGCAGATAAGTGGCCACAATGGCGGTAAAGCGGGTATTTAAACGCCACGTTGGCG 239  
181 ACGCCAGATAAGTGGCCACAATGGCGGTAAAGCGGGTATTTAAAGCGCACGTTGGCG 240  
240 GAGTTGTCATATGATGGGGGGTATTATGGTGAATTACTTTCCAGCCTGTTGGTTGTTA 299  
241 GAGTTGTCATATGATGGGGGGTATTATGGTGAATTACTTTCCAGCCTGTTGGTTGTTA 300  
300 GCGCGGGATGTCGCGCAGAGGGGGGTATATGCAAGGCAGATATCGATCAACGTAACC 359  
301 GCGCGGGATGTCGCGCAGAGGGGGGTATATGCAAGGCAGATATCGATCAACGTAACC 360  
360 GCGTTAACCGGACAGGCATTTCTGATGGTGGCTCGCGGAGTGGTTGACCGTCGAT 419  
361 GCGTTAACCGGACAGGCATTTCTGATGGTGGCTCGCGGAGTGGTTGACCGTCGAT 420  
420 AATCAGCGCATTTGAAAAAGCCGATCCCTGTTTCCATCCGGTAGAGCTGGCGTGTGT 479  
421 AATCAGCGCATTTGAAAAAGCCGATCCCTGTTTCCATCCGGTAGAGCTGGCGTGTGT 480  
480 AAGTTTACACACGACACAGATAAAGCGATTGAAGAAACGATTTGGCGGCGCGTTAGAT 539  
481 AAGTTTANCGNCGACACACAGATAAAGCGATTGAAGAAACGATTTGGCGGCGCGTTAGAT 540  
540 ACGGTGAGCAGCGCTACGCTAAACGTTTGGCCAGATGGGGGACGTGCTGAATTTGGCG 599  
541 ACGGTGAGCAGCGCTACGCTAAACGTTTGGCCAGATGGGGGACGTGCTGAATTTGGCG 600  
600 GCTTCTCTTTATTGCAAACTCTTTGCAACAGCAAGGAAACGTTGATTTTGGCCACTTT 659  
601 GCTTCTCTTTATTGCAAACTCTTTGCAACAGCAAGGAAACGTTGATTTTGGCCACTTT 660  
660 GCGGCCAATGAAGTTAAGTTAATGAAGGGAACAAAGTGACCTCAGTGGGCCACTG 719  
661 GCGGCCAATGAAGTTAAGTTAATGAAGGGAACAAAGTGACCTCAGTGGGCCACTG 720  
720 GCGTTATCATGACATTTGGTGAATCTTCTTATTACAAAACGACCAAGCCATGCCAGAG 779  
721 GCGTTATCATGACATTTGGTGAATCTTCTTATTACAAAACGACCAAGCCATGCCAGAG 780  
780 GTTCCCTGGCAACGGCTTAAAGGGGGAGAAATGGGTATCGTTATTGTCATTACATAAC 839  
781 GTTCCCTGGCAACGGCTTAAAGGGGGAGAAATGGGTATCGTTATTGTCATTACATAAC 840  
840 GCGCAATTTAATTTAATGGCAAAAACACCGTATATGCGCCCGTCAATAAGGGAACGCCATTA 899  
841 GCGCAATTTAATTTAATGGCAAAAACACCGTATATGCGCCCGTCAATAAGGGAACGCCATTA 900  
900 TTACAGCAGATAGATACGGCTTTAACCTTCACTGGAATGCTCAGGGGCAAAAGCTACCC 959  
901 TTACAGCAGATAGATACGGCTTTAACCTTCACTGGAATGCTCAGGGGCAAAAGCTACCC 960  
960 ATTTTCAGCCAAAACCGGGTCTTTGTTCTCGGTGGGCATGATACCAATATTGCCAAATTT 1019  
961 ATTTTCAGCCAAAACCGGGTCTTTGTTCTCGGTGGGCATGATACCAATATTGCCAAATTT 1020  
1020 GCGGGTATGCTGGAGCGGACTGGGAGCTTACCGGAGCAACCTGATTAATCTCCACCAAGT 1079  
1021 GCGGGTATGCTGGAGCGGACTGGGAGCTTACCGGAGCAACCTGATTAATCTCCACCAAGT 1080  
1080 GCGGGATTTGTTTGTGAACCTATGGCAAAATCCGGATAACCAACGACGCTACGTTGGGTG 1139



Db 841 GCACAGTTCACCTGATGCTAAACGCCATACATTCGACGCCACAAAGGACGCGGCTT 900  
Qy 900 TTACACGAGTACATACGCTTTAAACCTTCACTGATGCTCAGGGCABAAGTACCC 959  
Db 901 TTACAGCAATCATACCGCACTGACCTCTGCACTGACGCCAGGGGCAAAATCTGCCG 960  
Qy 960 ATTTCAGCCCAAAACCGGGTCTTTGTTCTCGGTGGGCATGATACCAATATTCCTCAATATT 1019  
Db 961 ATCTCGGCTCAGAACCGTGTATTTCTTCTGGGTGGCCACGACACAAATATTGCTAACATC 1020  
Qy 1020 GCGGTATGCTGGAGCCGACTGGCAGCTACCGAGCAACCTGATTAATCTCCACGAGT 1079  
Db 1021 GCGGTATGCTGGGCGCAGATTGGCAGTTACCGGAACAACCGGATTAACACCCCGGGC 1080  
Qy 1080 GCGGTATGCTTTTGAACATATGCGCAAAATCCGGAATAACACCGGCTACGTTGCGGTG 1139  
Db 1081 GCGGTCTGCTTTGAGCTGTGCGAGATTCGGACAATCATCAAGTTATGTGGCGGTT 1140  
Qy 1140 AAGATGTTCTACAAAGATGATCAGTTGCGTATGCGGAATAATGATCTGAAAT 1199  
Db 1141 AAGATGTTCTATCAGACCATGGATCAATTGCGTAACCGGAGAGCTGGATTTAAAGAAC 1200  
Qy 1200 AACCCAGCGGTATTATTTCCGTTGCGATGCTGCTGTTGTGAAATAACCGTGACGATAAG 1259  
Db 1201 AATCCCGCGCATCATCATGTCGCTGTGGCGGCTGCGAGATTAATGTCGACGATAA 1260  
Qy 1260 CTTTGGAGCTTGATACATTCGAAAGAAAGTGGCTAAGTAAATTAAGAACTGCTGCCAC 1319  
Db 1261 CTGTGCACTTGATATCTTTTCAAAAAAAGTAGCGAAATCATTTGAACCTGCTGTCAT 1320  
Qy 1320 ATCTAA 1325  
Db 1321 ATTAA 1326

RESULT 5  
ACF35789  
ID ACF35789 standard; DNA; 1281 BP.  
XX AC ACF35789;  
XX AC ACF35789;  
DT 06-NOV-2003 (first entry)  
DE Nov9X phytase polypeptide encoding DNA.  
XX Nov9X; thermotolerant; phytase; phosphate; animal food; phosphorous;  
KW gene; ds.  
XX Synthetic.  
OS  
FH Key Location/Qualifiers  
FT CDS 25..1263  
FT /\*tag= a  
FT /transl\_except= (pos; 25..27, aa: Met)  
FT /product= "Nov9X"  
XX W02003057247-A1.  
XX 17-JUL-2003.  
XX 30-DEC-2002; 2002WO-US041785.  
XX 28-DEC-2001; 2001US-0344523P.  
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX (DIVE-) DIVERSA CORP.  
XX Lanahan ML, Koepf E, Kretz K;  
XX WPI; 2003-598328/56.  
DR P-PSDB; ABR82310.  
XX Preparing thermotolerant phytase useful for preparing animal feed or

human food, by expressing in microbial host cell expression cassette comprising a promoter operably linked to a nucleic acid encoding the enzyme.  
XX Example 3; Page 77-78; 79pp; English.  
XX The invention relates to preparing a thermotolerant phytase (I), involving expressing in a microbial host cell an expression cassette comprising a promoter operably linked to a nucleic acid molecule encoding (I). The thermotolerant phytase has a specific activity of greater than 400 U/mg, 600 U/mg or 800 U/mg at pH 4.5 and 37plusC. The nucleic acid molecule encodes a fusion polypeptide comprising the thermotolerant phytase and a signal sequence which is operably linked to the thermotolerant phytase. The thermotolerant phytase is glycosylated and has a simulated gastric half-life of greater than 25 minutes at a pH greater than 2.0 and less than 4.0. (I) is useful for preparing animal feed, improving nutritive value of animal feed and human food, and for preparing human food. A heat-treated animal feed mixture comprising an inorganic phosphate at below 0.45% and (I) is useful for reducing feed conversion ratios or increasing weight gain of animals fed diets with inorganic phosphate at levels below 0.45%. It is also useful for decreasing phosphate levels in excreta from an animal. An animal feed composition comprising (I) is useful for decreasing the feed conversion ratio and increasing the weight gain of an animal. The composition is useful for minimizing inorganic dietary requirements of phosphorous in an animal, and for enhancing organic phosphate utilization from organic phosphorus sources in feed for an animal, where the feed increases the bioavailability of inorganic phosphorus in the feed to the animal. It is also useful for decreasing phosphate levels in excreta from an animal. In all the above mentioned conditions, the feed is a poultry or swine feed. The thermotolerant phytase has a half-life of about 30 minutes in the digestive tract of the animal. The present sequence represents a thermotolerant Nov9X phytase encoding DNA  
XX Sequence 1281 BP; 322 A; 295 C; 300 G; 364 T; 0 U; 0 Other;

Query Match 18.3%; Score 242; DB 8; Length 1281;  
Best Local Similarity 52.5%; Pred. No. 6.7e-68;  
Matches 628; Conservative 0; Mismatches 555; Indels 13; Gaps 4;  
Qy 93 ATCGGGCTATCTTTAGAACGTGTGGTATTTTTCAGTCGCCATGGTTCGCTCGCCGAC 152  
Db 39 ACCAGAAATTGAAGTTGGAACTCTGTGTCATTGTTCCAGACACGGTGTAGAGTCCAAC 98  
Qy 153 CAA-ACACGAGCAGCTTATGAATGATGTACGCCAGATAAGTGGCCACATGGCGGTA 211  
Db 99 TAAGCTACTCAGTTGATGCAAGATGTTACTCCAGATGCTTGGCTTACCTGGCTGTAA 158  
Qy 212 AGCGGGTATTTAAACGCCACGCTGTGGGAGTTGGTCACTGATGGGGGGTTTATGG 271  
Db 159 GTTGGGTGAATTGACTCCAGAGGTGGTGAATTGATTGCTTACTTGGGTCACTAGGAG 218  
Qy 272 TGATTACTTTCGACGCTTGTGTTTGT--AGCGCGGGATGTCGGCAGAGGGGGGT 328  
Db 219 ACAAGATTGGTGTGCTGATGGTTTGTTCGCAAGATGTTGTTCCAGATCTGTCAAGT 278  
Qy 329 ATATGCACAGGAGATATCGATCAACGTPACCGCTTAAACCGGACAGGATTTCTTGATGG 388  
Db 279 TGCTATCATTTGCTGATGTTGATGAAGAACTAGAAAGAACTGTTGAAGCTTCGCTGCCGG 338  
Qy 389 TGTGGCTCGGGGTGTTGACGCTGCATAATCAGGCCGATTTGAAAGAGACGATCC 448  
Db 339 TTTGGCCCCAGACTGTGCTATCACTGTTCACTCACTCACTCACTCACTCACTCACTCACT 398  
Qy 449 CCTGTTCCATCCGTTAGAGACTGGCGTGTGTAAGTTAGACAAACGACAAACAGATAAGC 508  
Db 399 ATTGTTCAACCATTTGAAGACTGGTGTCTGTCAATTGGATAACGCTTAACGTTACTGATGC 458  
Qy 509 GATTGAAGAACGATTGGCGGGCGCTTTAGATACGGTGAAGCCAGCGCTACGCTAAACCTTT 568  
Db 459 CATCTTGAAGAGAGCTGGTGGTTCTATCGCTGACTTCTGCTGCTACTACCAAACTGCCCTT 518



569	Qy	TGCCAGATGGGGACGTGCTGAAATTTTGGCGCTTCTCCTATTTCGAAATCTTTGCAACA	638
519	Db	CAGAGATTTGGAAAGAGTCTTTGAACCTTCCACAACTTAACCTGTGTTTGAAGAGAGAA	578
629	Qy	GCAAGGAAAAACGTGTGATTTTGCCCACTTTGGGCCAAATGAAGTTAAAGTTAATAAGA	688
579	Db	GCAAGCAGATCTTGTTCCCTTGACTCAAGCCTTGCCATCTGAATTBAGAGTCTCTGCTGA	638
689	Qy	AGGGAACAAAGTGAACCTCTAGTGGGCGCACTGGGGGTATCATTCGACATTTGGTGAAATCTTT	748
639	Db	TTG-----TGTCTCTCTGATGGTGTGTCTCTCTGGCTTCTATGTGTGACTGAAATCTTT	692
749	Qy	CTTATTACAAAAACGACAAAGCCATGCCAGAGTTGCCGCAACGGCTAAAGGGGGCGGA	808
693	Db	CTTTGTTCCAAAGCTCAAGGTATGCCAGAACAGGTTGGGTAGAATCACTGATTTCA	752
809	Qy	GAATTGGGTATCCTTATTGTCAATTACATAACGGCAATTTAAATTAATGCAAAAAACACC	868
753	Db	CCAATGAAACACCTTTGTGTCTTGCACAAACGCTCAATTCGAATTTGCTGCAGAGAACTCC	812
869	Qy	GTATATCGCCGTCATAAAGGGACGCATTTATACACAGATAGATACGGCTTTAACCTT	928
813	Db	AGAACTCGCTAGATCCAGAGCTACTCAATTTGTGGACTTTCGATCAAGACCCCTTTGAC---	869
929	Qy	TCAACTGGATGCTCAGGGGCAAAAGCTACCCATTTACGCCCAAAACGGGCTCTTGTTCCCT	988
870	Db	TCCACACCCACCAAGAAAGCAAGCTTACGGGTGTACCTGCCAACTCTCTGCTGTTTCAT	929
989	Qy	CGTGGGCATGATACCAATATTGCCAATATTGGGGTATGCTGGGAGCCCACTGGCAGCT	1048
930	Db	TGCCGCTCAGACTAACTATGGCTAACTTTGGGTGCTGCTTGGAAATGAACATGGACCTT	989
1049	Qy	ACCCGACCACTCGATAATACTCCACAGGTGGGGATTCGTTTTCGAACTATGGCAAAA	1108
990	Db	GCCAGGTCAACCAAGATAAACACTCCACAGGTGGTGAATTTGGTCTTCGAAGATGGCGTCG	1049
1109	Qy	TCCGGATAACCAACAGGCTTACGTTGGGTGAAGATGTTCTACCAAAAGATGGAATGTT	1168
1050	Db	ACTGTCTGATAACTCTCAATGGATTCAAAGTCTCCTTGGTCTTCCAAACCTTGCACAAAT	1109
1169	Qy	CGGTATATGCCAAAAATTTGGATCTGAAAAATTAACCCAGCGGTATTTATTTCCGTTGCAGT	1228
1110	Db	GAGAGCAAGACTCCATTTGTTCCTTGGAACACTCCACAGGTGAAGTCAAGTTGACCTTGGC	1169
1229	Qy	TGCTGGTTGTGAAAAATACCGGTGACGATAAGCTTTGCGAGCTTGATACATCCAAA	1284
1170	Db	TGTTTGTGAAGAAGAAACGCTCAAGGTATGTGTTCTTTGGCTGGTTTCACTCAAA	1225

## RESULT 6

AAC68299  
ID AAC68299 standard; DNA: 5421 BP.

XX AAC68299:

XX

XX  
XX  
XXXXXXX

XX  
DE  
SV40/AFPA pravima coating sequence:

KW Transgenic animal; salivary protein; phytase; phosphorus; animal growth;  
KW environmental pollution; piq: ds.

50  
XX

OS Escherichia coli.

3300

FN  
XX  
WU200008424/-AL.

PD 02-NOV-2000.  
XX

PF 20-APR-2000;  
yy

PR 23-APR-1999; 99US-0130508P.

[illegible]





Db 629 GTTCATTAACGAGGATACCATCGGAATCAAGGTGAGCGCGACTGTCTCTATTAA 688  
 QY 703 CCCTCAGTGGGCACTGCGCTTATCATCGACATTTGGGTGAATCTTCTATTACAAACG 762  
 Db 689 CC-----GTCGGGTAGCTCGCATCAATGCTGACCGAGATATTTCTCTGCAACAG 742  
 QY 763 CACAAGCCATGCGAGGTTGCTGCGCAACGGCTAAAGGGCGGAGATTTGGGTATCCT 822  
 Db 743 CACAGGGAATGCGGAGCGCGGTGGGAAGGATCACCGATTACACCGAGTGAACACCT 802  
 QY 823 TATTCTCATTAACAACGCGCAATTTAATTAATGGCAAAACACCGTATATCGCCGTC 882  
 Db 803 TGCTAAGTTTGCAHAACGCGCAATTTGATTGCTACACGCGCAGAGGTTCGCCGCA 862  
 QY 883 ATAAAGGAGCGCATATTATACAGCAGATAGATAGCGCTTTTAAACCTTCAACTGATGTC 942  
 Db 863 GCGCGCCACCCCGTTATTAGATTTGATCAAGACAGCGCTTGACGCCCATC---CACCGC 919  
 QY 943 AGGGCCAAAGCTACCCATTTACGCCCAACCGGCTCTTCTCTGGTGGGATGATA 1002  
 Db 920 AAAAAACAGCGTATGATGATTTACCATCTTACGCTGTTTATCGCCGACAGATA 979  
 QY 1003 CCAATATTGCCAATATTGGGGTATGCTGGAGCGGACTGGCAGCTACCCGAGCAACCTG 1062  
 Db 980 CTAATCTGCAATCTCGCGCGCACTGGAGCTCACTGGACGCTTCCCGGTGAGCGG 1039  
 QY 1063 ATATATCTCCAGCGTGGGATTTGTTTGAATGATGATGCAAAATCCGATACCAAC 1122  
 Db 1040 ATAAACGCGGATGATGATTTACCATCTTACGCTGTTTATCGCCGACAGATA 1099  
 QY 1123 AGCGTACGTTGCGGTGAAGATGTTTCTACCAAAACGATGATGATGCGTAAAT 1175  
 Db 1100 GCCAGTGGATTGAGTTTGGCTGGTCTTCCAGACTTTTACAGCAGATGCGTGT 1152

RESULT 8

AD25463  
 ID AAD25463 standard; DNA; 1901 BP.  
 AC AAD25463;  
 DT 26-MAR-2002 (first entry)  
 DE Escherichia coli appA phytase wild type DNA.  
 KW Bacterial phytase; K12 appA phytase; protease stability; anabolic;  
 KW gastrointestinal; nutritional value; feed treatment process; therapy;  
 KW thermal tolerance; growth performance; alcoholic drink; biopulping;  
 KW non-alcoholic drink; biobleaching; ds.  
 OS Escherichia coli.  
 FH Key  
 FT CDS  
 FT Location/Qualifiers  
 FT 188..1486  
 FT /tag= a  
 FT /product= "E. coli appA phytase protein"  
 PN WO200190333-A2.  
 PD 29-NOV-2001.  
 PE 24-MAY-2001; 2001WO-US017118.  
 XX 25-MAY-2000; 2000US-00580515.  
 PR (DIVE-) DIVERSA CORP.  
 XX Short JM, Kretz KA, Gray KA, Barton NR, Garrett JB, O'donoghue E;  
 PI WPI; 2002-083108/11.  
 DR P-PSDB; AAS15807.  
 XX

PT New bacterial phytase for e.g. improving the nutritional value of phytate  
 PT -containing foodstuffs and subsequently improving the growth performance  
 PT of an organism that consumes it, or in treating animal digestive systems.  
 XX Claim 42; Fig 7; 170pp; English.  
 PS  
 CC The patent discloses recombinant bacterial phytase from Escherichia coli  
 CC K12 appA phytase. The enzyme has phytase activity and improved thermal  
 CC tolerance when compared with wild-type phytase. It has improved protease  
 CC stability at low pH. The recombinant phytase is useful for improving the  
 CC nutritional value of phytate-containing foodstuffs and subsequently  
 CC improving the growth performance of an organism that consumes it, in  
 CC treating animal digestive systems, in feed treatment processes and for in  
 CC vitro purposes related to research, discovery and development. They are  
 CC also used for generating recombinant digestive system life forms, for  
 CC producing or manufacturing alcoholic and non-alcoholic drinks based on  
 CC the use of moulds, grains and/or plants, in biopulping and bio-bleaching  
 CC where a reduction in the use of environmentally harmful chemicals that  
 CC are traditionally used in the pulp and paper industry is desired and in  
 CC the reduction or possible elimination of the need for mineral  
 CC supplements, enzymes or therapeutic drugs for animals from the daily feed  
 CC thus increasing the amount calories and nutrients present in the feed.  
 CC The present sequence is a DNA encoding E. coli appA phytase wild type  
 CC protein  
 XX  
 SQ Sequence 1901 BP; 474 A; 499 C; 499 G; 428 T; 0 U; 1 Other;

Query Match 17.0%; Score 225.2; DB 6; Length 1901;  
 Best Local Similarity 53.2%; Pred. No. 2.7e-62;  
 Matches 571; Conservative 0; Mismatches 489; Indels 13; Gaps 4;  
 QY 107 TAGAAGCTGTGGTATTATTTGAGTCGCCATGCTGTTTCGCTCGCGCAGCAA-ACAACGCGAG 165  
 Db 276 TGGAAAGTGTGGTATTGTCAGTCGTCATGTTGGTGTGCTCCAAACGAGCCGACGAAC 335  
 QY 166 TTATGAATGATTTACGCGAGATAGTGGCCCAATATGGCGGTAAAGCGGGGTATTAA 225  
 Db 336 TGATGAGGATGTCAACCCAGACGATGCGCCCACTTGGCGGTAAAGCGGTGGCTGA 395  
 QY 226 CGCCAGCTGTGGTGGGAGTGTGCATTTGATGGGGGGTATTTATGTGTGATTAATTCGCA 285  
 Db 396 CACCGCGGTGTGGTGTGATTAATCGCTATCTCGACATTAACCAACGCCAGCTCTGTAG 455  
 QY 286 GCCTTGTGTTTGTAGCG---GCGGATGTCGCGCAGAGGGGGGGGTATATGCAACGCGAG 342  
 Db 456 CCGACGATGTCTGGCGAAGGAGGCTGCGCGCAGTCTGTCTGAGTCTGGGATTTATGCTG 515  
 QY 343 ATATCGATCAACCTTACCGCTTAACCGCAGCAGCATTTCTTGTATGTTGGTCCCGGGT 402  
 Db 516 ATGTGACGAGCGTACCCGTTAAACAGGCGAGCCTTCGCCCGCGGGCTGGACCTGACT 575  
 QY 403 GTGGTTTGACCGTGCATAATCAGGCGGATTTGAAAGAACCGATCCCTGTTTCCATCCGG 462  
 Db 576 GTGCANTAAACGTTACATCCAGGCAGATACGTCAGTCCCGATCCGTTATTTAATCCTC 635  
 QY 463 TAGAGACTGCGGTGTGTAGTAGAACAACGCAACAGATTAAGCGATTAAGAGACGAT 522  
 Db 636 TAAAAACTGGCGTTTGGCAACTGGAATAACGCGAACCTGACTGACCGGCTCTCAGCAGGG 695  
 QY 523 TGGCGCGCGGTTAGATACGTTGAGCGCGCTACGCTAAACCTTTTGGCCAGATCGGGG 582  
 Db 696 CAGGAGGTCAATTGCTGACTTTACCGGCATCGGCAACGCGGTTTCGCGAATCGAAC 755  
 QY 583 ACGTGCTGAATTTTGGGCTTCTCTTATTTGCAATCTTTGCAACAGCAGGAAACGT 642  
 Db 756 GGGTGTCTTAATTTTCCGCAATCAAACTGTGCTTTAAACGTGAGAACAGGACGAAGCT 815  
 QY 643 GTGATTTTGGCCACTTTGGCGCAATGAAGTTAACTTAATTAAGAGGACGACAAAAGTGA 702  
 Db 816 GTTCATTAACGAGGACATTACCATCGGAACCTCAGGTGAGCGCGCA-----CAATGTCT 869  
 QY 703 CCCTCAGTGGGCGCACTGGCGTTATCATTCGACATGCGGTGAAATCTTCTTTATCAAAACG 762



E. coli K12 appA gene encoding phytase.

Phytase; ds; gene; phytate; appA gene; animal feed; inorganic phosphate; digestion enhancement; transgenic; thermal tolerance; protease stability. Escherichia coli; strain K12.

Key Location/Qualifiers  
CDS 188..1486  
/\*tag= a  
/product= "Phytase"

US2002136754-A1.

26-SEP-2002.

24-MAY-2001; 2001US-00866379.

13-AUG-1997; 97US-00910798.  
01-MAR-1999; 99US-00259214.  
13-APR-1999; 99US-00291931.  
25-MAY-1999; 99US-00318528.  
25-MAY-2000; 2000US-00580515.

(SHORT J M.  
(KRETTZ K A.  
(GRAY K A.  
(BARTON N R.  
(GARR J B.  
(DONOHUE E.  
(MATHUR E J.

Short JM, Kretz KA, Gray KA, Barton NR, Garrett JB;  
O' Donoghue E, Mathur EJ;  
WPI; 2003-040002/03.  
P-FSDS; ADA19450.

Isolated Escherichia coli polynucleotide encoding a modified phytase enzyme, useful in the production of animal feed, for improving the nutritional value of phytate-containing feedstuff and for enhancing digestion in humans and animals.

Claim 42; Fig 7; 62pp; English.

The invention relates to an isolated Escherichia coli polynucleotide encoding a phytase enzyme appearing as ADA19450 and having amino acids modification from W68E, Q84W, Q84M, A35P, K97C, S168E, R180Y, N226C or Y277D. Also included the E. coli appA gene ADA19449 (or an oligonucleotide derived from it) or its mutant sequence ADA19452, expression vectors, host cells, a method of improving nutritional value of a phytate-containing feedstuff by contacting the phytate-containing feedstuff with a pure phytase enzyme mutated phytase (where the phytase enzyme catalyzes the liberation of inorganic phosphate from the phytate in the phytate-containing feedstuff), a method to produce an animal feed containing a microbial phytase (comprising culturing the plant cell, plant part or plant under conditions where the nucleotide sequence is expressed and converting the plant cells, plants or plant into a composition for animal feed), a feed composition for animals (comprising the plant seeds, plant cells, plant parts or plants in admixture with a phytate-containing feedstuff), a method to treat a human or an animal able to benefit from digestive enhancement by the activity of an exogenous phytase enzymes comprising administering to the human or animal the plant seed, plant cells, plant parts or plants of a transgenic plant which is modified to contain an expression system which expresses a nucleotide sequence encoding a phytase enzyme, a transgenic non-human organism whose genome comprising a heterologous nucleic acid sequence encoding a polypeptide having phytase activity. The phytase enzyme is useful for improving the nutritional value of phytate-containing feedstuff, in the production of animal feed and for enhancing digestion in humans and animals. The invented method improves thermal tolerance and protease stability. It also improves the feeding value of phytate rich ingredients. The present sequence represents the E. coli K12 appA gene encoding wild-type phytase.

XX	SQ	Sequence	1901 BP	474 A	499 C	499 G	428 T	0 U	1 Other
		Query Match	17.0%	Score	225.2	DB	7	Length	1901
		Best Local Similarity	53.2%	Pred. No.	2.7e-62				
		Matches	571	Conservative	0	Mismatches	489	Indels	13
								Gaps	4
QY	107	TAGAACTGTGTGTTATTTTGAAGTCGCGCATGGTGTTCGTCGCCGACAA-ACAACGACG	165						
DB	276	TGAAAGTGTGTGTATTTGTCAGTCGTATGTCGTGTGTCATCAACCAAGGCCACGCAAC	335						
QY	166	TTATGAATGATGTTACGCCAGATAAGTGGCCACAATGGCCGGTAAAGGGGGTATTTAA	225						
DB	336	TGATGAGGATGTCAACCCAGACGCAATGGCCAACTGGCCGGTAAACTGGTGGCTGA	395						
QY	226	CGCACTGTGTGGGAGTTGGTCACTGATGAGGGGGTTTTATGTTGATTAATTCTTCGCA	285						
DB	396	CACCGGNGGTGTGTAGCTAATCGCTATCTCGCAATTAACCAACGCCAGCTCTGGTAG	455						
QY	286	GCCTTCGTTGTTAGG- --GCGGGATGTCGCGCAGAGGGGGGTATATGACACGGCAG	342						
DB	456	CCAGCGATTGCTGGCGCAAAAGGGCTGCCCGCAGTCTGTGTCAGTTCGGATTATTGCTG	515						
QY	343	ATATCGATCAACGTACCCGCTTAACCGGACAGGCAATTTCTTGATGTTGGCTCCGGGGT	402						
DB	516	ATGTCAGCAGCGTATCCCGTAAACAGCGGAAGCTTCGCGCGCGGCTGGCACCTGACT	575						
QY	403	GTGGTTGACCGTGCATATTCAGCCGATTTGAAAGACCGATCCCTGTTCCATCCGG	462						
DB	576	GTGCAATAACCGTACATACCCAGCAGATACGTCCTGTCGCGATTCGTTATTAAATCCTC	635						
QY	463	TAGAGACTGCGGTGTGTAAGTTAGACAACGCAACAAACAGATAAAGCGATTGAAGAACGAT	522						
DB	636	TAAAACTGGGTTTGCCAACTGGATAACGCGAAGCTGACTGACGGGATCCTCAGCAGGG	695						
QY	523	TGGCGCGGCGTTAGATACGGTGAGCAGCGCTACGGCTAAACCTTTTGCCACGATGGGG	582						
DB	696	CAGGAGGTCAAATGCTGTACTTTACCGGGGATCGGCAACGGCGTTTCGCGAACTGGAAC	755						
QY	583	ACGTGCTGAATTTGCGGCTTCTCCTTATTGCAAACTTTTGCAACGACGAAGGAAAAACGT	642						
DB	756	GGTGTCTTATTTCCGCAATCAAACTGTGCTCTTAAACGTGAGAAACAGACGAAAGCT	815						
QY	643	GTGATTTTGCCACTTTGCGGCCAAATGAAGTTAACGTTTAAATAAGAGGACAAAAAGTGA	702						
DB	816	GTTCATTAACGCGAGCAATTACATCGGAACTCAAGGTGAGCGCGCA-----CAATGTCT	869						
QY	703	CCCTCAGTGGGCCACTGCGGTATCATCGCATTTGGGTGAATCTCTTATTACAAAACG	762						
DB	870	CATTAAACGGTGGGTGAAGCCTGCATCAATGTGACGAGATATTTCCTCTGCAACAG	929						
QY	763	CACAAGCATGCGCAGAGTTTCCTTGGCAACGCTTAAAGGGGGGAGAAATGGGTATCCT	822						
DB	930	CACAGGGAATGCGGAGCGGGGTGGGGAAGGATCACCGATTACACAGTGGGAACACCT	989						
QY	823	TATGTTCATTACATAAGCGCAATTTAATTTATGCGCAAAAACACGTATATGCCCGTC	882						
DB	990	TGCTTAAGTTTGCATAACCGGCAATTTATTGTTCTAACGCAACGCCAGAGGTTGCCGCA	1049						
QY	883	ATAAGGGAAGCCATTATTACAGCAGATAGATACGGCTTTAAACCTTTCAACTGGATGCTC	942						
DB	1050	GCCGCGCAACCCGTTATTAGATTGATCAGACAGCGTTGACGCCCATC---CACCGC	1106						
QY	943	AGGGGCAAAAGCTACCAATTTACGCCCAAAACCGGGTCTGTCTCTCGTGGGATGATATA	1002						
DB	1107	AAAAACAGGCGTATGGTGTGCAATTACCCACTTCAGTGTCTGTTTATCGCCGACACGATA	1166						
QY	1003	CCAATATTGCCAAATTATGCGGGTATGCTGGGAGCCGACTGCGACGTACCCGAGCAACTG	1062						
DB	1167	CTAATCTGGCAAACTCTGGCGGCGCATGAGAGTCACTGGACGTTCCCGGTGACGGG	1226						
QY	1063	ATAATACTCCACAGGTGGGGATTTGTTTTTGAACATATGGCAAAATCCGGATAACCACT	1122						

Db 1227 ATAAACAGCCGCCAGGTGGTGAACGTGGTGGTTTGAACCGCTGGCGGTAAAGCATACACA 1286

QY 1123 AGCGGTACGTTCCGGTGAAGATGTTCTACCAAAACGATGGATCAGTTGCGTAAT 1175

Db 1287 GCCAGTGGATTACAGTTCCTGGCTGGTCTCTCCAGACTTTACACAGATGGGTGAT 1339

RESULT 11

ADC87744

ID ID ADC87744 standard; DNA; 1901 BP.

AC AC

XX ADC87744;

XX

XX 01-JAN-2004 (first entry)

XX

XX DNA encoding Escherichia coli appA phytase.

XX

XX Phytase; food supplement; enzyme delivery matrix; soybean meal;

KW thermotolerance; thermostability; kernel; phytate;

KW myo-inositol-hexaphosphate; inositol; inorganic phosphate;

KW thermotolerant; feed value; digestion; gene; ds; appA.

XX

XX Escherichia coli.

XX

XX Location/Qualifiers

Key 188..1486

FT CDS /tag= a

FT /product= "AppA phytase"

FT /transl\_except= (pos:401..403,aa:Arg)

FT

XX

XX US2003103959-A1.

XX

XX

XX 05-JUN-2003.

XX

XX 24-MAY-2002; 2002US-00156660.

XX

XX 13-AUG-1997; 97US-00910798.

PR 01-MAR-1999; 99US-00259214.

PR 13-APR-1999; 99US-00291931.

PR 25-MAY-1999; 99US-00318528.

PR 25-MAY-2000; 2000US-00580515.

PR 24-MAY-2001; 2001US-00866379.

XX

PA (DIVE-) DIVERSA CORP.

XX

XX Short JM, Kretz K, Gray KA, Barton NR, Garrett JB, O'donoghue B;

PI Mathur EJ;

PI

XX

DR WPI; 2003-787039/74.

DR P-PSDB; ADC87745.

XX

XX New nucleic acid encoding a polypeptide having phytase activity, useful in improving the feeding value of phytate rich ingredients or as an aid in phytate digestion.

XX

XX Example 1; SEQ ID NO 3; 113pp; English.

XX

CC The invention discloses a new isolated or recombinant nucleic acid which encodes a polypeptide having a phytase activity. Also claimed is a nucleic acid probe, an amplification primer sequence pair, an expression cassette comprising the nucleic acid, a vector comprising the nucleic acid, a transgenic non-human animal or plant, or its seed, comprising the nucleic acid, an antisense oligonucleotide, inhibiting the translation of a phytase message in a cell, a heterodimer comprising the polypeptide and a second domain, an array comprising immobilised polypeptide or nucleic acid, a hybridoma comprising an antibody that specifically binds to the polypeptide, a food supplement for an animal, an edible enzyme delivery matrix, an edible pellet comprising a granule edible carrier and the polypeptide, a feed composition, a soybean meal, isolating or identifying the polypeptide, making an anti-phytase antibody, producing a recombinant polypeptide, determining whether a compound binds to the polypeptide, identifying a modulator, whole cell engineering of new or modified phenotypes by using real-time metabolic flux analysis, increasing

CC thermotolerance or thermostability of the phytase polypeptide, increasing  
CC the resistance of the phytase polypeptide to enzymatic inactivation in a  
CC digestive system and processing of corn and sorghum kernels. The phytase  
CC activity comprises catalysis of phytate (myo-inositol-hexaphosphate) to  
CC inositol and inorganic phosphate and is thermotolerant. The nucleic acid  
CC is useful in improving the feeding value of phytate rich ingredients or  
CC as an aid in phytate digestion. The sequence presented is the DNA  
CC encoding the *Escherichia coli* appA phytase.

XX SQ Sequence 1901 BP; 474 A; 499 C; 499 G; 428 T; 0 U; 1 Other;

Query Match 17.0%; Score 225.2; DB 9; Length 1901;  
Best Local Similarity 53.2%; Pred. No. 2.7e-62;  
Matches 571; Conservative 0; Mismatches 489; Indels 13; Gaps 4;

Qy 107 TAGAACGTGGTATTGAGTCGCCATGGTGTGCTCGCCGACCAA-ACAAACGACG 165  
Db 276 TGGAAAGTGGTATTGATGTCAGTCGTGTCGTGTCGTCCTCCAAACCAAGCCACGCAAC 335  
Qy 166 TTATGATGATGTTACCCAGATGAAGTGGCCACATGCGCCGCTAAGAGCGGGTATTAA 225  
Db 336 TGATGAGGATGTCACCCAGAGCATGGCCACCTGGCCGGTAAACTGGGTGGCTGA 395  
Qy 226 CGCCACGTGGTGGAGTGGTTCACATGATGGGGGGTTTTATGGTGAATTACTTCGCA 285  
Db 396 CACCGCGGTGGTGGATGATGCTATCTCGACATTAACCAACGCGCTCGTAG 455  
Qy 286 GCCTGGTGGTGGAGG---GCGGATGTCGGCAGAGGGGGGTATATGCACAGCGAG 342  
Db 456 CGGACGATGCTGGCGAAAGGGCTGCCCGAGTCTGTGTCAGGTGCGGATATTGCTG 515  
Qy 343 ATATCGATCAACGTACCGGCTTAAACCGGACAGGCAATTTCTTGATGGTGGTCCGGGGT 402  
Db 516 ATGTCGACGAGCGTACCCGTAACAGCGCAAGCCCTTCGCGCGGCTGGCACCTGACT 575  
Qy 403 GTGGTTTGCAGTCGTCATTAACAGCCGATTTGAAAGACGATCCCTGTTTCATCCGG 462  
Db 576 GTGCAATTAACCGTATACACCCAGGACATGCTGCTGCTGCTGCTGCTGCTGCTGCT 635  
Qy 463 TAGAGACTGGGCTGTGAAGTTAGACAAACGACAAACAGATTAAGCGATTGAAGACGAT 522  
Db 636 TAAAACTGGGTTTGGCACTGATTAACGGGACGTGACTGACGGGATCTCAGCAGGG 695  
Qy 523 TGGCGCGGCTTATAGATACGGTGAAGCAGCGCTACGCTAAACCTTTTGGCCAGATGGGG 582  
Db 696 CAGGAGGGTCAATTTGCTGACTTTACCGGCGATCGGCAACGGGGTTCGCGAACTGGAAC 755  
Qy 583 ACGTGTGTAATTTTGGCGCTTCTCTTATTGCAATCTTTGCAACGACGAAGGAAACGT 642  
Db 756 GGGTGTATTTTCCGCAATCAAACTGTGCTTAAACGTGAGAAACAGACGAAAGCT 815  
Qy 643 GTGATTTGGCCACTTTGCGCCCAATGAAGTTAACTTAAGTAAAGAGGACAAAAGTGA 702  
Db 816 GTTCATTAACGACGCGCATTAACATCGGAATCTCAAGGTGAGCGCGA-----CAATGCT 869  
Qy 703 CCCTCAGTGGCCACTGGCTTATCATCGACATTTGGTGAATCTTCTTATTACAAACG 762  
Db 870 CATTAACGCGGTGAAGCTTCGATCAATGCTGACGGAGATTTCTCTCTGCAACAG 929  
Qy 763 CACAAGCCATGCCAGAGTTTCCTGGCAACGGCTTAAAGGGGCGGAGAAATGGGTATCCCT 822  
Db 930 CACAGGGAATCCGAGGCGGGTGGGGAAGATCAACCGATTACACAGCTGGAACACCT 989  
Qy 823 TATTGTCTATCATAAACGCGCAATTTAATTAATGGCAAAACACCGTATATCGCCGCT 882  
Db 990 TGCTAAGTTTGCATAACGCGCAATTTTATTTTGTACAAACGCAACCGGAGGTTGCCGCA 1049  
Qy 883 ATAAAGGACGCCATTATTACAGCAGATAGATAGCGGTTTAAACCTTTCACTGGATGCTC 942  
Db 1050 GCGCGCCACCCGTTATTAGATTGATCAGACAGCGTTGACGCCCATC---CACCGC 1106  
Qy 943 AGGGGCAAAAGCTACCCATTTACGCCCAAACCGGGTCTTGTCTCTCGTGGGCAATGATA 1002

Db 1107 AAAAACACGGCGTATGGTGTGACATTAACCACTTCAGTGTCTGTTTATCGCCGACACGATA 1166  
Qy 1003 CCAATATTGCCAATAATTTCGGGGTATGCTGGGAGCGGACTGGCAGTACCCGAGCAACCTG 1062  
Db 1167 CTAATCTCGCAAAATCTCGCGCGGCGACTGGAGCTCAACTGGACGCTTCCCGGTGAGCGG 1226  
Qy 1063 ATATACTCCACCAGGTGGGGGATTTGTTTGAATGCTGCAAAATCCGGATACCAACC 1122  
Db 1227 ATACACACCGCCGCGAGGTGGAACCTGGTGTGTTGAACGCTGGCGTGGGTAAAGCGATAACA 1286  
Qy 1123 AGCGCTAGCTGGCGGTGAAGATGTTCTACCAAAAGATGGATCAGTTGGGTAAAT 1175  
Db 1287 GCCAGTGGATTCAGGTTTCGCTGCTTCCAGACTTTACAGCAGATGCGTAT 1339

RESULT 12  
ABK12514

ID ABK12514 standard; DNA; 1299 BP.

XX AC ABK12514;

XX DT 05-JUN-2002 (first entry)

XX DE DNA encoding phytase associated protein.

XX KW Phytase; gene; ds.

XX OS Unidentified.

XX FH Key Location/Qualifiers

FT CDS 1..1299

FT /\*tag= a

FT /product= "Phytase associated protein"

XX PN KR99086028-A.

XX PD 15-DEC-1999.

XX PF 25-MAY-1998; 98KR-00018810.

XX PR 25-MAY-1998; 98KR-00018810.

XX PA (WOJ-) WOJIN CO LTD.

XX PI Bae HD, Forceburgh CW, Goloben S, Cheng KJ;

XX DR WPI: 2000-645078/62.

XX P-PSDB; AAU7775.

XX PI Novel phytase gene, recombinant phytase and usage thereof.

XX PS Claim 1; Fig 2; 10pp; Korean.

XX CC The invention relates to a novel phytase gene, a recombinant phytase gene and their uses. This sequence encodes a phytase associated protein, described in the invention

XX SQ Sequence 1299 BP; 317 A; 345 C; 355 G; 282 T; 0 U; 0 Other;

Query Match 17.0%; Score 224.6; DB 3; Length 1299;  
Best Local Similarity 53.2%; Pred. No. 3.4e-62;

Matches 571; Conservative 0; Mismatches 489; Indels 13; Gaps 4;

Qy 107 TAGAACGTGGTATTGAGTTCGCCATGAGTGGCTCGCCGACCAA-ACAAACGACG 165

Db 89 TGGAAAGTGGTATTGATGTCAGTCTGTCATGGTGGTGTGCTCAACCAAGCGCACGCAAC 148

Qy 166 TTATGAATGATGTTACGCCAGATAGTGGCCACATAGTGGCCGCTGTAAGCGGGTATTTAA 225

Db 149 TGATGAGGATGTCACCCAGACGATGGCCACCTGGCCGCTAAGACTGGGTGGCTGA 208

Qy 226 CCGCAGCTGGTGGGAGTTGTCATGATGGGGGGTTTTATGGTGAATTTCTTCGCA 285



Db 209 CACCGCGCGGTGGTGAAGTAATCGGCTATCTCGGACATTACCAACGCCAGCGTCTGGTAG 268  
 QY 285 GCGTTGGTTTGTAGCG---GCGGATGTCGGCAGAGGGGGGTATATGCACAGGCAG 342  
 Db 269 CCGACGGATTGCTGGCGAAGAGGCTGCCGACGTCTGGTCAAGTCCGGATTATGCTG 328  
 QY 343 ATATCGATCAACGTACCGCTTAAACCGACAGGCAATTTCTGATGCTGGCTCGGGGT 402  
 Db 329 ATGTCGACGAGCGTACCGCTTAAACAGGGAAGCTTCCGCGCGGGCTGGACCTGACT 388  
 QY 403 GTGGTTTGACCGTGCATATCAGCGCGATTTGAAAAGACCGATCCCTGTTTCCATCCGG 462  
 Db 389 GTGCAATACCGTACATACCCAGGAGATACGTCAGTCCCGATCCGTTATTTAATCTC 448  
 QY 463 TAGAGCTGCGTGTGTAGTATAGCAACGCAACAGATTAACCGATTGAGAACGAT 522  
 Db 449 TAAAACCTGCGCTTGGCCAACTGGATAACGCGAAGCTGACGCGATCTCTACGACGG 508  
 QY 523 TGGCGGCGCGTTAGATACGGTGAGCCAGCGCTAGCTTAAACCTTTGCGCCAGATGGGG 582  
 Db 509 CAGGAGGTCAATGCTGACTTACCGGCGATCGGCAACAGCGGTTTCGCGACTGGAC 568  
 QY 583 AGTGTGTAATTTGGCGCTTCTCTTATGCAATCTTTGCAACAGCAAGCAAAAGCT 642  
 Db 569 GGGTGTCTTAATTTCCGCAATCAAACTTGTGCTTAAACGCTGAGAAACAGGACGAACT 628  
 QY 643 GTGATTTTCCCACTTTGGCGCAATGAAGTTAAAGTTAATAAGAGGCAAAAGTGA 702  
 Db 629 GTTCATTAACGAGCATTAACATCGAATCAAGTGAAGCGCGA-----CATGCT 682  
 QY 703 CCTCAGTGGCGCATGGGCTTATCATCGACATATGGGTGGAATCTTCTTATTAACAAAG 762  
 Db 683 CATTAACCGGTGGTGAAGCTCGCATCAATCTGACGGAGATATTTCTCTCGCAACAG 742  
 QY 763 CACAAGCCATGCCAGAGGTGCTGCAACGCTTAAAGGGCGGAGATTTGGGTATCCT 822  
 Db 743 CACAGGATGCGAGCGCGGGTGGGAGGATCACCGATTACACCACTGGAACACT 802  
 QY 823 TATGTCATTAACAGCGCAATTTAATTAATGCAAAACACCGTATATCGCCGTC 882  
 Db 803 TGCTAAGTTGCAATACCGGCAATTTATTTGCTACAGCGCAGCGAGGTTCGCCGCA 862  
 QY 883 ATAAAGGAGCGCATTAATTAACAGCATAGATACGCTTTAACCTTCACTGATGTC 942  
 Db 863 GCGCGCCACCGCTTATGATTTATGATTTATCAAGACAGCTTGACGCCCATC---CACGCG 919  
 QY 943 AGGGGCAAAAGCTTACCCATTTCAGCCCAAAACCGGCTCTTTGCTCGGTGGGCGATGATA 1002  
 Db 920 AAAAAACAGGCGTATGTTGACATTACCCACTTCAGTCTGTTTATCGCGGACACGATA 979  
 QY 1003 CCAATATTGCAATATGCGGTATCTGGAGCCGACTGGCAGCTACCGGAGCACTG 1062  
 Db 980 CTAATCTGGCAATCTCGCGGCGCACTGGAGCTCAACTGGAGCGCTTCCCGGTGACCGG 1039  
 QY 1063 ATAATACTCCACAGGTGGGGGATTTGGTTTTTGAACATATGGCAAAATCCGGATAACCCAC 1122  
 Db 1040 ATAAACGCGCGAGGTGGTGAATCGTGTGTTGAACGCTGGCTCGGCTAAGCATAA 1099  
 QY 1123 AGCGTACGTGGGTGAAGATTTCTACCAACGATGATCGTTGCGTAAAT 1175  
 Db 1100 GCAGTGGATTACAGTTTCGCTGCTTCCAGACTTTACAGCAGATGCGTGTAT 1152

RESULT 13

ACAL9297  
 ID ACA19297 standard; DNA; 1299 BP.

XX AC

XX ACA19297;

XX 19-JUN-2003 (first entry)

XX Prokaryotic essential gene #954.

XX

KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
 KW drug design; gene.  
 OS Escherichia coli.  
 XX WO200277183-A2.  
 XX 03-OCT-2002.  
 XX 21-MAR-2002; 2002WO-US009107.  
 XX 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362899P.  
 XX (ELIT-) ELITRA PHARM INC.  
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX WPI; 2003-029926/02.  
 DR P-PSDB; ABU15427.  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 for homologous nucleic acids required for cellular proliferation to  
 isolate candidate molecules for rational drug discovery programs.  
 XX Claim 14; SEQ ID NO 7167; 1766pp; English.  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 the 6213 antisense sequences given in the specification where expression  
 of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 encoding a polypeptide whose expression is inhibited by the antisense  
 nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 polypeptide or its fragment whose expression is inhibited by the  
 antisense nucleic acid; (4) an antibody capable of specifically binding  
 the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 the gene product or that has an activity against a biological pathway  
 required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 pathway in which a proliferation-required gene or its gene product lies  
 or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX Sequence 1299 BP; 317 A; 345 C; 355 G; 282 T; 0 U; 0 Other;

Query Match 17.0%; Score 224.6; DB 7; Length 1299;

Best Local Similarity 53.2%; Pred. No. 3.4e-62;

Matches 571; Conservative 0; Mismatches 489; Indels 13; Gaps 4;

QY 107 TAGAACGTGTGTTATTTGATCGCCATGTGTGCTCGCCACCAA-ACAAAGCAGC 165

Db 89 TGGAAAGTGTGTGTTATTTGATCGTATGTGTGCTGCTCAACCAAGCCACGCAAC 148



QY	166	TTATGAATGATGTTACGCCAGATAAGTGGCCACAAATGCCCGTAAAGCGGGTATTAA	225
DB	149	TGATGACGAGATGTCAACCCAGACGATGCCCAAACCTGGCCGGTAAAACTGGGTTGGCTGA	208
QY	226	CGCCACGTGTGGCGAGATTGGTTCACATTGATGGGGGGTTTATATGCTGATTTAC	285
DB	209	CACCGCGGGTGGTGAAGCTAATCGCTATCTCGACATTAACCAACGCCAGCGTCTGGTAG	268
QY	286	GCCTTGGTTTGTATAGC--CGGGAGATCTCGGCAGAGGGGGGGGTATATGCA	342
DB	269	CCGACGGATGCTGGCGGAAAAAGGGCTCCCGCGCATCTGTCAGGTCGCGATATTGCTG	328
QY	343	ATATCGATCAACGTACCCCGCTTAACCGCACAGGCATTTCTTTGATGCTGTGCTCCGGGT	402
DB	329	ATGTCGACGAGGTACCCGTAAACACAGCGAAGCCTTCGCCCGGGCTGGCACTGACT	388
QY	403	GTGGTTGACGTGTCATTAATCAGGGCGGATTTGAAAAAGACCGATCCCTCTGTTCATCGG	462
DB	389	GTGCAATTAACCGTACATACCAGGCAGATAGCTCCAGTCCCGATCCGTTATTTAATCCTC	448
QY	463	TAGAGACTGGGCTGTGAAGTTAGACAACGACACAAACAGATAAAGCGATTGCAAGAACGAT	522
DB	449	TAAAACTGGCGTTTGCCAACTGGATAACGGGAACGTGACTGACGCGATCTCTAGCAGGG	508
QY	523	TGGCGGGCCGTTAGATACGGTGAAGCGTACGCTAAACCTTTTGGCCAGATGGGGG	582
DB	509	CAGGAGGTCATTAATGTCTGACTTTACCGGGCATTCGGCAAACGGCGGTTTCGGAACTGGAA	568
QY	583	ACGTGCTCAATTTTTCGGGCTCTCCCTTATGCAAACTTTTGCACACAGCAAGGAAAAACGT	642
DB	569	GGTGTCTTAATTTTCGGCAATCAAACTTGTGCTTAAACGTGAGNAACAGACGAAAGCT	628
QY	643	GTGATTTTGCCCACTTTTCGGGGCAATGAAGTTAAGTTAATAAAGAGGGGACAAAGTGA	702
DB	629	GTTCATTAACCGAGGCATTAACCATCGGAACTCAAGGTGAGCGCCGA-----CAATGTCT	682
QY	703	CCCTCAGTGGGCCACTGGCGCTTATCATCGCAATTGGGTGAAATCTTCTTATTACAAACG	762
DB	683	CATTAAACGGTGGGTAGCCTCGCATCAAATGCTGACGAGATATTCTCTCTGCAACAG	742
QY	763	CACAAAGCATGCGAGAGTTGCTGCGCAACGGCTAAAGGGGGCGGAGAAATTTGGGTATCCT	822
DB	743	CACAGGGAATCCGAGAGCCGGGTGGGAGAGGATCACCGATTCACACCAAGTGGAAACCT	802
QY	823	TATTGTCAATTACATAACCGCAATTAATTTAATGGCAAAACACCGTATATCCCGCTC	882
DB	803	TGCTAAGTTTGCATTAACCGCAATTTATTGCTTACACGCGACGCGAGAGTTGCCCGCA	862
QY	883	ATAAAGGAGCGCAATTATACAGCAGATAGATACGGCTTTAACCCCTTCAAATGGAATGCT	942
DB	863	CCCGCGCACCCCGTTATTAGATTGATCAAGACAGCGTTGACGCCCCCATC---CACCGC	919
QY	943	AGGGGCAAAAGTACCCATTTACGCCCAAAACCGGGTCTTGTCTCTCGTGGCGCATGATA	1002
DB	920	AAAAACAGCGGTATGGTGTGACATTAACCACTTCAGTGTCTGTTTATCGCGGACAGGATA	979
QY	1003	CCAATATTGCCAATATTGCGGGTATGCTGGGAGCCGACTGGCAGGTACCCGAGCAACTG	1062
DB	980	CTAATCTGGCAAACTCCCGCGCGCACTTGGAGCTCAACTGACACGCTTCCCGGTACGCGG	1039
QY	1063	ATAATACTCCACAGTGGGGATTTGGTTTTTGAACATATGCCAAAAATCCGGAATACCAAC	1122
DB	1040	ATAACACGCCAGTGTGATCTGGTTGTTTGAACGCTGGCTCGGTAAACGATACGATTAACA	1099
QY	1123	AGCGGTACGTTGCGGTGAAGATGTTCTACAAACGATGGATCAGTTGGCTAAT	1175
DB	1100	GCCAGTGGATTCAGGTTTCGTTGGTCTTCAGACTTTACAGCATATGGTGTAT	1152

RESULT 14  
AAC68298  
ID AAC6  
XX

AC	AAC68298;	
XX		
DT	15-SEP-2003 (revised)	
DT	20-FEB-2001 (first entry)	
XX		
DE	R15/APPA plasmid coding sequence.	
XX		
KW	Transgenic animal; salivary protein; phytase; phosphorus; animal growth;	
KW	environmental pollution; pig; ds.	
XX		
OS	Rattus sp.	
OS	Escherichia coli.	
OS	Chimeric.	
XX		
FN	WO200064247-A1.	
XX		
PD	02-NOV-2000.	
XX		
PF	20-APR-2000; 2000WO-CA000430.	
XX		
PR	23-APR-1999; 99US-0130508P.	
XX		
PA	(UYGU-) UNIV GUELPH.	
XX		
PI	Foreberg CW, Golovan S, Phillips JP;	
XX		
DR	WPI; 2000-687245/67.	
DR	P-PSDB; AAB36261.	
XX		
PT	Transgenic non-human animal for gastrointestinal tract specific	
PT	expression of a protein, preferably phytase, comprises a nucleic acid	
PT	sequence including a heterologous transgene construct encoding the	
PT	protein.	
XX		
PS	Claim 14; Fig 21; 152pp; English.	
XX		
CC	The present invention provides transgenic animals which produce desired	
CC	proteins, in this case pigs which expresses phytase in the salivary	
CC	gland. Low phytase production levels result in phytate in the diet being	
CC	excreted and causing phosphorus contamination in water, as well as	
CC	reducing the growth of animals. The invention provides a number of	
CC	transgenes containing the E. coli APPA phytase coding sequence. (Updated	
CC	on 15-SEP-2003 to standardise OS field)	
XX		
SQ	Sequence 3470 BP; 1065 A; 721 C; 735 G; 949 T; 0 U; 0 Other;	
	Query Match 17.08; Score 224.6; DB 3; Length 3470;	
	Best Local Similarity 53.2; Pred. No. 6.1e-62;	
	Matches 571; Conservative 0; Mismatches 489; Indels 13; Gaps 4;	
QY	107 TAGAACGTGTGGTTATTTCGAGTCGCCATGTGTGTGCTCGCGACCAA-ACAACGCAGC 165	
Db	1999 TGGAAAGTGGTGATTGTCACTCGTCATGTGTGCGTGTCTCCAACCAAGGCCACGCAAC 1958	
QY	166 TTATGATGATGTTACCCCGATGAATGGGCCACAATGGCCGGTAAAGCGGGGTATTAA 225	
Db	1959 TGATGCAGGATGTCAACCCAGACGCATGGCCAACTGGCCGGTAAAACTGGGTGGCTGA 2018	
QY	226 CGCCACGTGGTCGGAGTTGGTCACATGATGGGGGGTTTATGTGTGATTACTTCGCA 285	
Db	2019 CACCGCGGTGGTGACTAATCGCCTATCTCGACATTAACACGCCAGCGTCTGGTAG 2078	
QY	286 GCCTTGTGTTGTAGCG---CGCGGATGTCCGCGACAGGGGGGGGTATATGCACAGCAG 342	
Db	2079 CCGACGGATGTCTGCCGAAAAGGGCTGCCCGCAGTCTGTGTCAAGTCGGCATTTATGCTG 2138	
QY	343 ATATCGATCAACGTACCCTGTTAACCGACAGACGATTTCTTGATGTGTGCTCCGGGT 402	
Db	2139 ATGTGCAGAGCGTACCCTGGTAACACGCGAGCCTTCGCCCGCGGTGCGACTGACT 2198	
QY	403 GTGTTTTGACCGTGCATATCAATCAGCCCGGATTGTAAAGACCGATCCCTGTTCCATCCGG 462	
Db	2199 GTGCAATAACCGTACATACCCAGGACGATAGTCCAGTCCCGATCGGTATTATTAATCCTC 2258	

463	QY	TAGAGACTGGCGTGTGAATAGTTAGACAACGCAAAACAGATAAAGGATTGAAGAACGAT	522
2259	Db	TAAAAACTGGCGTTTCCCACTGGATAAACGGACCGTGAATGACGCGATCTCTCAGCAGGG	2318
523	QY	TGGCGGGCGGCTTTAGATACGGTGAGCAGCGCTACGCTAAACCTTTTGCACCAGATGGGG	582
2319	Db	CAGGAGGGTCAATTGCTGACTTTACCGGGCATCGGCAACCGGGCTTTCGCGAACTGGAAC	2378
583	QY	AGTGCTGTAATTTGGCGCTTCTCCTATTATGCAATCTTTGCAACAGCAGAGGAAAACGT	642
2379	Db	GGGTGCTTAATTTTTCGCAATCAAACTGTGCTTAAACGTGAGAAACAGGACGAAAGCT	2438
643	QY	GTGATTTGCCCACTTTGGCGCAATGAAGTTAAACGTTAATAAGAAAGGACAAAAGTGA	702
2439	Db	GTTCAATAACGCAGGCATTACCATCGGAATCAAGGTGAGCGCGCA-----CAATGTCT	2492
703	QY	CCCTCAGTGGCCACTGGCGTTATCATCGACATTGGGTGAAATCTTTATTACAAAAAG	762
2493	Db	CATTACCGGTGCGTAAGCCTCGCATCAATGCTGACGGAGATATTTCTCTGTCACAAG	2552
763	QY	CACAAGCCATGCCAGAGGTTGCTGCGCAACGGCTAAAGGGGCGGAGAAATGGGTATCT	822
2553	Db	CACAGGGAATGCCGAGCGGGGTGGGAAGGATCACCGATTCAACACAGTGGAAACACT	2612
823	QY	TATTGTCATTACATAACGCGCAATTTAATTAATGCAAAAAACACCGTATATCGCCCGTC	882
2613	Db	TGCTAAGTTGCAATAACGCGCAATTTATTGCTCAACGCGCGCAGAGGTGCCCGCA	2672
883	QY	ATAAAGGAGCGCCATTATTACAGCAGATAGATACGGCTTTAACCCTTCAACTGGATGCTC	942
2673	Db	GCGCGCCACCCGGTTATTAGATTGATCAAGACAGCGTTGACGCCCATC---CACCGC	2729
943	QY	AGGGGCAAAAGCTACCCATTTACGCCCAAAACCGGGTCTGTTCCTCGGTGGGCAATGATA	1002
2730	Db	AAAAACAGCGGTATGGTGCAATTACCCACTTCAGTCTGTTTATTCGCGGNACAGATA	2789
1003	QY	CCAATATTGCAATTTGCGGGTATGGTGGAACCGACTGGGACGCTACCGCAGAACCTTG	1062
2790	Db	CTAATCTGGCAAACTCTCGCGCGCGCACTGGAGCTCAACTGGACGCTTCCCGGTACGCGG	2849
1063	QY	ATAATACTCCACAGGTGGGGGATGGTTTTTCAACTATGGCAAAATCCGGATAACCAAC	1122
2850	Db	ATPACAGCGCCAGGTGGTGAATGGTGTGTGAACTGGTGTGTGAACTGGCTGCGCTAAGCGATAACA	2909
1123	QY	AGCGCTACGTTGCGGTGAAGATGTTCTTCAAAAAGATGGATCAAGTTCAGTTCGGTAAT	1175
2910	Db	GCCAGTGGATTCAGGTTTCGCTGGCTTCCAGACTTTACAGAGATGCGGTAT	2962

RESULT 15	
AAC68296	
ID	AAC68296 standard; DNA; 4060 BP.
XX	
XX	
AC	AAC68296;
XX	
DT	15-SEP-2003 (revised)
DT	20-FEB-2001 (first entry)

XX DE R15/APPA plasmid coding sequence.

Transgenic animal; salivary protein; phytase; phosphorus; animal growth;  
 environmental pollution; pig; ds.

XX	Rattus sp.
OS	Escherichia coli.
OS	Chimeric.

XX PN WO200064247-A1.

XX PD 02-NOV-2000.

XX  
DE 20-APP-2000-2000WO-CAB000430

XX	23-APR-1999;	99US-0130508P.	
PR	(UYGU-) UNIV	GUELPH.	
XX	Forsberg CW,	Golovan S,	Phillips JP;
XX	WPI; 2000-687245/67.		
XX	P-PSDB; AAB36259.		
XX	Transgenic non-human animal for gastrointestinal tract specific		
PT	expression of a protein, preferably phytase, comprises a nucleic acid		
PT	sequence including a heterologous transgene construct encoding the		
PT	protein.		
XX	Claim 14; Fig 19; 152pp; English.		
XX	The present invention provides transgenic animals which produce desired		
CC	proteins, in this case pigs which expresses phytase in the salivary		
CC	gland. Low phytase production levels result in phytate in the diet being		
CC	excreted and causing phosphorus contamination in water, as well as		
CC	reducing the growth of animals. The invention provides a number of		
CC	transgenes containing the E. coli APPA phytase coding sequence. (Updated		
CC	on 15-SRP-2003 to standardise OS field)		
XX	Sequence 4060 BP; 1257 A; 814 C; 843 G; 1146 T; 0 U; 0 Other;		
XX	Query Match	17.0%; Score 224.6; DB 3; Length 4060;	
XX	Best Local Similarity	53.2%; Pred. No. 6.6e-62;	
XX	Matches 571; Conservative	0; Mismatches 489; Indels 13; Gaps 4;	
QY	107	TAGAAGCTGTGGTATTATTGAGTCGCCCATGTGTCGCTCGCCGACCAA-ACAACGCGAGC	165
DB	1899	TGGAAAGTGTGTGATTGTCAGTCGTGTCATGGTGTGCTGTCTCCAAACGAGCCACGCAAC	1958
QY	166	TTATGAAATGATGTTACGCCAGATAGTGGCCACATGGCCGGTAAAGCGGGGTATTAA	225
DB	1959	TGATGACAGATGTCACCCAGACGATGCGCAACCTGGCCGGTAAACCTGGGTGGCTGA	2018
QY	226	CGCCACGTGTGCGGAGTTGGTCACATGATGGGGGGGTTTATGTGTGTTACTTTCGCA	285
DB	2019	CACCGCGGTGGTGAGCTAACTCGCTCTCGACATTACCAACGCCGCTCTGTAG	2078
QY	286	GCCTTGGTGTGTAGG---GCGGATGTCGCGACAGGGGGGTATATGCACAGCGAG	342
DB	2079	CCGACGGATGCTTGGCGGAAAAAGGGCTGCCCGCAGTCTGGTCAAGTCGCGATTTTCCTG	2138
QY	343	ATATCATCAACGTACCCGCTTAACCGGACAGAGCATTTCTTGATGTGTGGCTCCGGGGT	402
DB	2139	ATGTGACGAGCGTACCGTAAACAGGCGAAGCCTTCGCCGCCGGCTGCACCTGACT	2198
QY	403	GTGGTTTGACCGTGCATATATCAGGCCGATTTGAAAAGACCGATCCCTGTTCCATCCGG	462
DB	2199	GTCCAAATAACCGTACATACCACGCGGAGATACGTCCAGTCCCGATTCGGTATTTAATCCTC	2258
QY	463	TAGAGACTGGCGTGTGTAGTTAGACAACGCCACAAACAGATAAAGCGATTGAAGAAACGAT	522
DB	2259	TAAAACTGGCTTGGCCAACTGATAACCGGAACGTGACTGACGCGATCCTCAGCAGGG	2318
QY	523	TGGGGCGCGCTTAGATACGGTGACCGCGCTACGCTAAACCTTTTGGCCAGNATGGGG	582
DB	2319	CAGGAGGGTCAATTGCTGACTTTACCGGGGATCGGCAACGGCGTTTCGGGAACCTGGAAC	2378
QY	583	ACGTGCTGAATTTTGGGGCTTCTCCTTATTGCAAAATCTTTTGCAACAGCAAGGAAAAACGT	642
DB	2379	GGGTGCTTAATTTTCGCAATCAAACTTGTCCTTAAACGTGAGAAACAGGACGAAAGCT	2438
QY	643	GTGATTTTGGCCACTTTTGGGGCCAAATGAAGTTAAAGTAAAGAGGAGCAAAAGTGA	702
DB	2439	GTTTCATTAAACGACAGGCAATTAACATCGGAACCTCAAGGTGACGCCGA-----CAATGCT	2492
QY	703	CCCTCAGTGGGCACTGGCGTTATATCGACATTTGGGTGAAATCTTCTTATTACAAACG	762

Db	2493	CATTAAACCGGTGCGGTAAAGCCTCGCATCAATGCTGACGGAGATATTTCTCTGCAACAAG	2552
Qy	763	CACAAGCCATGCCAGAGGTTGCCCTGGCAACGGCTAAAAGGGGGAGAAATGGGTATCCT	822
Db	2553	CACAGGGATGCCGGAGCCGGGTGGGAGGATCACCGATTACACCAAGTGGACACCT	2612
Qy	823	TATTGTCATTACATAACGGCGAATTTAATTTAATGGCAAAAACACCGTATATCGCCGTC	882
Db	2613	TGCTAAGTTTGCATAACCGCAATTTTATTGCTACAAACGACGCCAGAGGTTGCCCGCA	2672
Qy	883	ATAAAGGAGCCCATTTATACAGCAGATAGATACGGCTTTAAACCTTCAACTGGATGCTC	942
Db	2673	GCCGCGCACCCCGTTATTAGATTGATCAGACAGCGTTGACGCCCCATC---CACCGC	2729
Qy	943	AGGGGCAAAAGCTACCCATTTACGCCCAAAACCGGGTCTTGTCTCGTGGGATGATA	1002
Db	2730	AAAAACAGCGCTATGGTGTGACATTAACCACTTCAGTGTCTTTTATCGCCGGACACGATA	2789
Qy	1003	CCATATTGCCAATATTGCGGTATGCTGGAGCCGACTGCGAGCTACCCGAGCAACCTG	1062
Db	2790	CTAATCTGGCAATCTCGCGCGGACTTGGAGCTCAACTGACCGCTTCCCGTCAAGCGG	2849
Qy	1063	ATAATCTCCACAGGTGGGGATTGGTTTTTGAACATAATGGCAAAATCCGGATAACCAAC	1122
Db	2850	ATAACAGCCGCCAGGTGGTGAACCTGGTGTGTTGAACGCTGGCGTGAAGCGATAACA	2909
Qy	1123	AGCGCTAGTTGCGGTGAAGATGTTCTACCAACGATGGATCAGTTGGTAAAT	1175
Db	2910	GCCAGTGAATTCAGGTTTCGCTGGTCTTCCAGACTTTACAGCAGATGCGTAT	2962

Search completed: May 2, 2004, 22:50:08  
Job time : 382.238 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.  
OM nucleic - nucleic search, using sw model  
Run on: May 2, 2004, 22:22:03 ; Search time 72.6484 Seconds  
(without alignments)  
10121.490 Million cell updates/sec  
Title: US-10-021-723B-3  
Perfect score: 1325  
Sequence: 1 atgtcagattagaaatcg.....aacctgcctgcacatctaa 1325  
Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 682709 seqs, 277475446 residues  
Total number of hits satisfying chosen parameters: 1365418  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries  
Database : Issued Patents NA:\*  
1: /cgn2\_6/prodata/2/ina/5A-COMB.seq:\*  
2: /cgn2\_6/prodata/2/ina/5B-COMB.seq:\*  
3: /cgn2\_6/prodata/2/ina/6A-COMB.seq:\*  
4: /cgn2\_6/prodata/2/ina/6B-COMB.seq:\*  
5: /cgn2\_6/prodata/2/ina/6C-COMB.seq:\*  
6: /cgn2\_6/prodata/2/ina/backfiles1.seq:\*  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	218.8	16.5	1323	3	US-09-259-214-1
2	218.8	16.5	1323	3	US-08-318-528-1
3	218.8	16.5	1323	3	US-09-291-931-1
4	218.2	16.5	1489	4	US-09-540-149A-9
5	143.6	10.8	1272	2	US-08-910-798-1
6	53.8	4.1	1266	4	US-09-489-039A-341
7	48	3.6	1569	4	US-09-489-039A-6330
8	48	3.6	5975	1	US-08-920-812-23
9	48	3.6	5975	1	US-08-920-827-23
10	48	3.6	5975	1	US-08-921-177-23
11	48	3.6	5975	1	US-08-362-577C-23
12	48	3.6	5975	2	US-08-920-828-23
13	37	2.8	7218	1	US-08-232-463-14
14	35	2.6	1602	4	US-09-252-991A-4062
15	35	2.6	1773	4	US-09-252-991A-4093
16	34.2	2.6	402	4	US-08-921-976-17355
17	34.2	2.6	5185	4	US-09-976-594-640
18	33.8	2.6	9775	4	US-09-647-540A-1
19	33.8	2.6	9775	4	US-09-647-540A-3
20	33.8	2.6	9775	4	US-09-647-540A-5
21	33.8	2.6	9775	4	US-09-647-540A-7
22	33.8	2.6	9775	4	US-10-119-600-1
23	33.8	2.6	9775	4	US-10-119-600-3
24	33.8	2.6	9775	4	US-10-119-600-5
25	33.8	2.6	9775	4	US-10-119-600-7
26	33.8	2.6	9775	4	US-10-119-651-1
27	33.8	2.6	9775	4	US-10-119-651-3

ALIGNMENTS

RESULT 1  
US-09-259-214-1  
; Sequence 1, Application US/09259214A  
; Patent No. 6110719  
; GENERAL INFORMATION:  
; APPLICANT: Kretz, Keith  
; TITLE OF INVENTION: NOVEL PHYTASE  
; FILE REFERENCE: DIVER1370-1  
; CURRENT APPLICATION NUMBER: US/09/259,214A  
; EARLIER FILING DATE: 1999-03-01  
; EARLIER APPLICATION NUMBER: 08/910,798  
; EARLIER FILING DATE: 1997-08-13  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 1323  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1320)  
; NAME/KEY: misc feature  
; LOCATION: (1)...(1323)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-259-214-1

Query Match	16.5%	Score	218.8	DB	3	Length	1323
Best Local Similarity	52.8%	Pred. No.	1.4e-63				
Matches	567	Conservative	0	Mismatches	493	Indels	13
Gaps	4						
QY	107	TAGAACGTGTGGTATTATTTTGGTTCGTCATGTCCTGCCGACCAA-ACAACGACGC	165				
DB	89	TGGAAAGTGTGGTATTATTTTGGTTCGTCATGTCCTGCCGACCAA-ACAACGACGC	148				
QY	166	TTATGATGATGTTACGCCAGATAAGTGGCCACAAATGCGCGGTAAGCGGGGTATTAA	225				
DB	149	TGATGCGAGATGTACCCGACGACATGCGCGGTAAGCGGGGTATTAA	208				
QY	226	CGCCACGTGTGGTTCGTCATGTCCTGCCGACCAA-ACAACGACGC	285				
DB	209	CACCGCGGTGTGGTTCGTCATGTCCTGCCGACCAA-ACAACGACGC	268				
QY	286	CCCTTGGTTTCTTAGCG--CGCGGATGTCCGCGAGAGGGGGGTATATGCACAGGAG	342				
DB	269	CCGACGATTTCTGCGGAAAAAGGGCTGCCCGAGTCTGGTTCAGGTTCGCGATTATTGCTG	328				
QY	343	ATATCGATCAAGTACCGCTTAACCGACAGGAGGATTTCTTGATGGTGTGCTCGGGGT	402				
DB	329	ATGTCGACGAGTACCGCTTAACCGACAGGAGGATTTCTTGATGGTGTGCTCGGGGT	388				

Sequence 5, Appli  
Sequence 7, Appli  
Sequence 10334, A  
Sequence 10535, A  
Sequence 10639, A  
Sequence 18033, A  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 7, Appli  
Sequence 2311, Ap  
Sequence 46, Appl  
Sequence 46, Appl  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 23, Appl  
Sequence 59, Appl



QY 1003 CCAATATTCCTATTTGGGGTATGCTGGAGCCGACTGCTACCCGAGCACTG 1062  
DB 980 CTAATCTGCGAAATCTCGCGCGGCACTGAGCTCAACTGGACGCTTCCCGTCAAGCGG 1039  
QY 1063 ATAATACTCCACAGTGGGGATTTGGTTTGAATATGCGAAATCCGGATAACCCACC 1122  
DB 1040 ATAACACGCGCGAGTGGTGAATGTTTGAACGCTGGCGTCCGCTAAGCGATAACA 1099  
QY 1123 AGCGTACCTTGGGCTGAAGATGTTCTACCAACGATGATCAGTTGGTAAAT 1175  
DB 1100 GCCAGTGGATTCAGGTTTCGCTGGCTTCCAGACTTTACAGCAGATGGGTGAT 1152

RESULT 3  
US-09-291-931-1  
; Sequence 1, Application US/09291931A  
; Patent No. 6190897  
; GENERAL INFORMATION:  
; APPLICANT: Kretz, Keith  
; TITLE OF INVENTION: NOVEL PHYTASE  
; FILE REFERENCE: 09010/029003  
; CURRENT APPLICATION NUMBER: US/09/291,931A  
; CURRENT FILING DATE: 1999-04-13  
; EARLIER APPLICATION NUMBER: 08/910,798  
; EARLIER FILING DATE: 1997-08-13  
; EARLIER APPLICATION NUMBER: 09/259,214  
; EARLIER FILING DATE: 1999-03-01  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 1323  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1320)  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(1323)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-291-931-1

Query Match 16.5%; Score 218.8; DB 3; Length 1323;  
Best Local Similarity 52.8%; Pred. No. 1.4e-63;  
Matches 567; Conservative 0; Mismatches 493; Indels 13; Gaps 4;

QY 107 TAGAAGTGTGGTATTTAGTGGCCATGCTGGTTCGCTCGCGACCAA-ACAAGCGAGC 165  
DB 89 TGGAAAGTGTGGTATTTAGTGGCCATGCTGGTTCGCTCGCGACCAA-ACAAGCGAGC 148  
QY 166 TTATGATGATGTTAGCCGAGTAACTGGCCCAATGGCCGCTTAAAGCGGGTATTAA 225  
DB 149 TGATGAGGATGTCACCCAGAGCGATGGCCAACTGGCCGCTTAAAGCGGGTATTAA 208  
QY 226 CGCCAGTGTGGGAGTGTGTCACATGATGGGGGGTTTTATGGTGATTACTTTTCGCA 285  
DB 209 CACCGGCGGTGGTGGTAACTGCTATCTCGGACATACCAAGCCAGGCTGGTAG 268  
QY 286 GCCTTGGTTTGTAGCG---GCGGATGTCGCGACAGGGGGGGGTATATGACAGGCGAG 342  
DB 269 CCGACGGATGCTGGCGGAAAGGGCTGGCCGAGTCTGGTCAAGTGGCGATTTGCTG 328  
QY 343 ATATCGATCAACGTACCCGCTTAAACCGGACAGGCAATTTCTTGATGGTGGTTCGCGGT 402  
DB 329 ATGTCAGAGCGTACCCGTTAAACAGGCGAAGCCCTTCGCGCGGCTGGCACTGACT 388  
QY 403 GTGGTTGACCGTGCTATATCAGCGCGATTTGAAAGACCGATCCCTGTTCCATCCGG 462  
DB 389 GTGCAATAACCGTACATACCCAGCGAGATACGTCAGTCCCGATCCGTTATTATTCCTC 448  
QY 463 TAGAGACTGGCGTGTGATGTAGCAACGCAACAGATAAGCGATTTGAAGACGAT 522  
DB 449 TAAAACTGGCGTTTCCCACTGGATAACGGAACGTGACTGACGGGATCCTCAGCAGG 508

QY 523 TGGCGGGCGCTTAGATACGGTGGAGCCGCTACGCTAAACCTTTTCCAGATGGGG 582  
DB 509 CAGAGGGCTCAATGCTGACTTTACCGGGCATGGCAACCGGGTTTCGGAAGTGAAC 558  
QY 583 ACGTGCTGAATTTTGGCGCTTCTCTTTATTTGCAAACTTTTGCACAGCAAGGAAAAAGT 642  
DB 569 GGGTGTCTTAATTTTCCGCAATCAAACTTGTGCTTAAACGCTGAGAAACAGGACGAAAGCT 628  
QY 643 GTGATTTTGGCCACTTTGGCGGCCAATGAAGTTAACGTTTAAATGAAGGAGGACAAAGTGA 702  
DB 629 GTTCATTAACGCGAGGCATTTACCATCGGAATCAAGGTGAGCGCGCA-----CAATGTCT 682  
QY 703 CCCTCAGTGGGCCACTGGCGTTATCATCGACATTTGGTGAATCTTCTTTATTAACAAACG 762  
DB 683 CATTAACCGGTGGGTAAAGCTCGCATCAATGCTGACGGAGATATTTCTCTCTGCAACAG 742  
QY 763 CACHAGCCATGCCAGAGGTTGCTGGCAACGGCTAAAGGGGGGAGAAATGGGTATCTCT 822  
DB 743 CACAGGGAATGCCGGAGCCGGGGTGGGGAAGGATCAACCGANTCACACAGTGAACACCT 802  
QY 823 TATTGTCATTACATAACGCGCAATTTTAAATTTGATGCAAAACACCGTATATCGCCCGTC 882  
DB 803 TGCTAAGTTTGCATTAACGCGCAATTTTATTTGCTACACGCGCAGAGGTTTCCCGCA 862  
QY 883 ATAAAGGAGCGCAATTTATACAGCAGATAGATAGGCTTTTAAACCTTCAACTGATGCTC 942  
DB 863 GCGCGCCACCGCTTATTGGATTTTGATCATGGCAGCGTTGACGCGCCCATCCAC---CGC 919  
QY 943 AGGGGCAAAAGTACCCATTTTACGCCCAAAACCGGGTCTTGTCTCGGTGGGCATGATA 1002  
DB 920 AAAACAGGGGTATGGTGTGACATTACCCACTTCACTACTGTTTATTTGCCGACAGATA 979  
QY 1003 CCAATATTGCAATATTTCGGGTATGCTGGAGCCGACTGGCAGTACCCGAGCAACCTG 1062  
DB 980 CTAATCTGGCAAAATCTCGCGGGCGCACTGGAGTCAACTGGAGCTTCCCGGTGAGCGG 1039  
QY 1063 ATATACTCCACAGGTGGGGATTTGTTTGAATATGCGCAAAATTCGCGATAACCAAC 1122  
DB 1040 ATAAACGCGCCAGGTGGTGAATGTTGTTGAACGCTGGCTGGCTGAGCGATTAACA 1099  
QY 1123 AGCCTACGTTGGGTGAAGATGTTTACCAACGATGATGATGATGCGTAAAT 1175  
DB 1100 GCCAGTGGATTCAGGTTTCGCTGGTCTTCCAGACTTTACAGCAGATGGGTGAT 1152

RESULT 4  
US-09-540-149A-9  
; Sequence 9, Application US/09540149A  
; Patent No. 6511699  
; GENERAL INFORMATION:  
; APPLICANT: Lei, Xingen  
; TITLE OF INVENTION: ENZYMES WITH IMPROVED PHYTASE ACTIVITY  
; FILE REFERENCE: 19603/2791  
; CURRENT APPLICATION NUMBER: US/09/540,149A  
; CURRENT FILING DATE: 2000-03-31  
; PRIOR APPLICATION NUMBER: 60/127,032  
; PRIOR FILING DATE: 1999-03-31  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 9  
; LENGTH: 1489  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
US-09-540-149A-9

Query Match 16.5%; Score 218.2; DB 4; Length 1489;  
Best Local Similarity 52.8%; Pred. No. 2.4e-63;  
Matches 567; Conservative 0; Mismatches 493; Indels 13; Gaps 4;  
QY 107 TAGAAGTGTGGTATTTAGTGGCCATGCTGGTTCGCTCGCGACCAA-ACAAGCGAGC 165  
DB 270 TGGAAAGTGTGGTATTTAGTGGCCATGCTGGTTCGCTCGCGACCAA-ACAAGCGAGC 329

QY 166 TTATGAATGATGTTAGCGCAGATAAGTGGCCACAAATGCGCGTAAAGCGGGTATTAA 225  
DB 330 TGATGACGATGTACCCGACGATGCGCAACCTGCGCGTAAAGTGGTGGTGA 389  
QY 226 GCGCAGTGTGCGGAGTGTGTCATATGATGGGGGGTATTATGGTGAATTACCTTCGCA 285  
DB 390 CACCAAGCGGTGGTGGTGTATGCTATCTCGACATTAACAAGCCAGCGTGTGGTG 449  
QY 286 GCTTGGTGTGTTAGCG--GGGGATGTCGGCAGAGGGGGGGTATATCCACAGGAG 342  
DB 450 CCGACGGATGTGCGGAAAAGGGCTGCCGCGACCTGTCAGGTGCGGATTTATGCTG 509  
QY 343 ATATCGATCAACGTACCGCTTAACCGGACAGGATTTCTTGATGGTGTGGCTCGGGGT 402  
DB 510 ATGTCGACGAGTACCGGTAAACAGCGGAGCTTCGCGCGGGCTGGACCTGACT 569  
QY 403 GTGGTTGACCGTGCATATACAGCGCGATTTGAAAAGACGATCCCTGTTCCATCGG 462  
DB 570 GTGCAATAACCGTACATACCGAGGAGATAGTCCAGTCCGATCCGTTATTATATCCT 629  
QY 463 TAGAGACTGGGTGTGTAAGTTAGCAACGACGACAAACAGATAAAGCGGATTAAGAAGAT 522  
DB 630 TAAAAAATCTGGGTTTCCACCTGATGTAACGGGACGTGACGCGATCTCAGAGGG 689  
QY 523 TGGGCGGCGGTAGATACGAGTACGCGCTAGCTTAACCTTTTGGCCAGATGGGG 582  
DB 690 CAGGAGGTCATCTGCTGACTTTACCGGCGCATCGCAAAAGCGCTTTCGCAACTGGAAC 749  
QY 583 ACGTCTGAATTTTGGCGCTCTCTTATTCGAATCTTTGCAACGACGAGGAAAGCT 642  
DB 750 GGTGCTTAATTTTCCCAATTAACCTTGCTTAACTGCTTAACTGAGAAACGAGCAAGCT 809  
QY 643 GTGATTTTCCCATCTTTCGCGCAATGAAGTTAAAGTTAATAAAGAGGACAAAGTGA 702  
DB 810 GTTCATTAACGAGCACTTACCATCGGAATCAAGGTGAGCGCGA-----CAATGTT 863  
QY 703 CCTCAGTGGCCACTGCTGTATCATGCAATTTGGTGAATCTTCTTATTAACAAAG 762  
DB 864 CATTAAACGGTTCGGTAAAGCTTCGATCAATGCTGACGGAATATTTCTCTGCAACAG 923  
QY 763 CACAAAGCCATGCGAGGTTGCTGCGCAACGCTTAAAGGGGGCGGAGAAATGGGTATCCT 822  
DB 924 CACAGGGAATGCGGAGCGGGTGGGAGGATCACTGATTCACACGAGTGAACACCT 983  
QY 823 TATGTCATATCATTAACGCGCAATTTAATTAATGGGAAACACCGTATATCGCGTC 882  
DB 984 TGCTAAGTTTGCATAACGCGCAATTTATTTACTACAAACGCAACGAGGTTCCCGCA 1043  
QY 883 ATAAAGGAGCGCATTTATACAGCAGATAGATACGGCTTTAACCTTCACTGATGCTC 942  
DB 1044 GTCGCGCCACCCGTTATGATTTGATTCATGCGAGCGTTGACGCCCATC---CACGC 1100  
QY 943 AGGGGCAAAAGTACCCATTTTCAGCCCAAAACCGGGTCTTGTCTCGGTGGGATGATA 1002  
DB 1101 AAAAAACAGCGGTATGGTGTGACATTTACCCACTTCAGTGTCTGTTATTTCGCGACACGATA 1150  
QY 1003 CCAATATTGCCAATATGCGGTATGCTGGAGCGGACTGCGAGCTACCCGAGCAACTG 1062  
DB 1151 CTAACTGGAATCTCGCGCGCAGCTGAGCTCACTGACGCTTCAGGTGAGCGG 1220  
QY 1063 ATAACTCCACAGGTGGGGAATTTGTTTGAATATGCAAAATCCCGATACCAACC 1122  
DB 1221 ATAAACACGCGCGAGTGGTGAATCTGTTTGAACGCTGCGTGGCTGAAGCGATAACA 1280  
QY 1123 AGCGCTACGTTGCGGTGAAGATGTTCTACCAACGATGATGATGCTGCTAAT 1175  
DB 1281 GCCAGTGGATTCAGGTTTCGCTGCTTCCAGACTTTACAGCAGATGCGTAT 1333

RESULT 5  
US-08-910-798-1  
; Sequence 1, Application US/08910798

; Patent No. 5876997  
; GENERAL INFORMATION:  
; APPLICANT: KRETZ  
; TITLE OF INVENTION: NOVEL PHYTASE  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: California  
; COUNTRY: US  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/910,798  
; FILING DATE: August 13, 1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: HAILE PH.D., LISA A.  
; REGISTRATION NUMBER: 38,347  
; REFERENCE/DOCKET NUMBER: 09010/029001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619/678-5070  
; TELEFAX: 619/678-5099  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1272 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; IMMEDIATE SOURCE:  
; CLONE: PHYTASE  
; FEATURE:  
; NAME/KEY:  
; LOCATION:  
; US-08-910-798-1

Query Match 10.8%; Score 143.6; DB 2; Length 1272;  
Best Local Similarity 50.2%; Pred. No. 4.5e-38;  
Matches 539; Conservative 0; Mismatches 470; Indels 64; Gaps 5;  
QY 107 TAGAAGTGTGGTATTATTTAGTGGCCAGATAAGTGGCCCAATGGCCGTAAAAGCGGGTATTTAA 165  
DB 89 TGAAGTGTGGTGAATGTGTCAGTGTGCTGCTCAACCAAGGCGCAGCAAC 148  
QY 166 TTATGAATGATGTTACGCCAGATAAGTGGCCCAATGGCCGTAAAAGCGGGTATTTAA 225  
DB 149 TGATGAGATGTCAACCCAGAGCATGGCCAACTGGCCGTAAAACCTGGTGGCTGA 208  
QY 226 CGCCACGTGGTGGGAGTTGGTCAATTGATGGGGGGTATTTATGGTGAATTAATTCGCA 285  
DB 209 CACCGCGGTGGTGAATTAATCGCTATCTCGACATTTACCAACGCCAGCGTCTGGTAG 268  
QY 286 GCCTTGTGTTTAGCG---GGGGATGTCGCGAGAGGGGGGTATATGCACAGGAG 342  
DB 269 CCGACGATGTGCGGAAAAAGGGCTGCCAGTCTGTCAGTGGCGATTAATTCGTC 328  
QY 343 ATATCGATCAACGTACCCGCTTAACCGGACAGGCAATTTCTTGAATGGTGGTCCGGGT 402  
DB 329 ATGTCAGCAGCGTACCCGTAACCAAGGCGAGCA----- 363  
QY 403 GTGTTTGGCCGTGCATATACGCGCGATTTGAAAAGACCGATCCCTGTTCCATCCGG 462  
DB 364 -----GATGTCAGTCCGATCCCGATCCGTTATTATTCCTC 397  
QY 463 TAGAGACTGGCGTGTGTAAGTTAGACACGCGCAACAGATAAAGCGATTAAGAACGAT 522  
DB 398 TAAAAACTGCGCTTTCGCACTGGAFAACGCGAACTGACTGACGCGATCCTCAGCAGGG 457



QY 523 TGGGCGGCGCTTAGATACGGTACGCGGCTACGCTAAACCTTTTGGCCAGATGGGG 582  
Db 458 CAGGAGGGTCAATGTGCTGACTTACCGGCGATCGGCAACCGCGCTTTCGCGAATCGAAC 517  
QY 583 AGCTGCTGAATTTTGGCGGCTTCTCCCTATTGCAATCTTTGCAACAGCAAGGAAAAAGCT 642  
Db 518 GGGTGTCTTAATTTTCGCAATCAAACTTGTGCTTAAACGCTGAGAAACAGGAGCAAGCT 577  
QY 643 GTGATTTTCCGCACTTTGGGCGCAATGAAGTTAACTTAATTAAGAGGGAACAAAGTGA 702  
Db 578 GTTCATTAACGCGAGGCAATACCATCGGAATCAAGGTGAGCGCCAGAC-----AATGTCT 631  
QY 703 CCTCAGTGGGCGCACTGGCGTTATCATCGACATTTGGGTGAATCTTCTTATTACAAAAG 762  
Db 632 CATTAACCGGTGGGTAAAGCTCGCATCACTGCTGACGAGATATTTCTCTGCAACAG 691  
QY 763 CACAAGCCATGCGAGAGTTGCTGTCACAGGCTAAAGGGGCGGAGAAATGGGTATCCT 822  
Db 692 CACAGGGAATGCGGAGCGGGTGGGGAAGGATCACCGATTCAACAGTGGAAACCT 751  
QY 823 TATTGTCAATACATACGCGCAATTTAATTTAATGCAAAAACCGGTATATCGCCGTC 882  
Db 752 TGTAAATTTGCAATACGCGCAATTTTATTGCTACAGCGCGGAGAGTTGCCCCGA 811  
QY 883 ATAAAGGAGCGCATTTATACAGCATAGATACGCTTTAAACCTTCAACTGGATGCTC 942  
Db 812 GCGCGCCACCCCGTTATTGGATTGATCATGCGAGCGTTGACGCCCATCCAC---CGC 868  
QY 943 AGGGGCAAAAGCTACCCATTTCAGCCCAAAACGGGTCTGTTCCTGCTGGGATGATA 1002  
Db 869 AAAACAGCGGTATGCTGACATTAACCCACTTCACTGATGTTATTGCGGACACGATA 928  
QY 1003 CCAATATTGCCAATATTGCGGGTATGCTGGGAGCGCACTGGCAGCTACCCGAGCAACCTG 1062  
Db 929 CTATCTGCGCAATCTCGCGCGGCACTGGAGCTCACTGGAGCTTCCCGTACGCGG 988  
QY 1063 ATAACTCTCACAGTGGGGAATGTTGTTTGAATATGGAATAATCGGAATAACCCACC 1122  
Db 989 ATAAACGCGCGCAGTGGTGAATGTTGTTGAACTGTTGAACTGGGCTGGGCTAAGCGATA 1048  
QY 1123 AGCGCTACCTTGGGTTGAAGATGTTCTTACCAACGATGATCAGTTGGTAT 1175  
Db 1049 GCCAGTGGATTGAGTTTCGCTGGCTTCCAGACTTTACAGCAGATGGGTAT 1101

RESULT 6  
US-09-489-039A-341  
; Sequence 341, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489, 039A  
; PRIOR FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 341  
; LENGTH: 1266  
; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-341

Query Match 4.1%; Score 53.8; DB 4; Length 1266;  
Best Local Similarity 49.0%; Pred. No. 1.3e-07;  
Matches 230; Conservative 0; Mismatches 232; Indels 7; Gaps 3;

QY 57 ATTGGCTGCTATTACTGCGCGGTAGCGCGGCGGCTATCTTTAGACGCT 116  
Db 66 ATCCGCTGGCGTCTAGCGCGAGACAGGCGCGCGGAGGGGTATCAGCTCAGAGT 125

QY 117 GGTATTATTTAGTGGCCATGGTGTTCGCTGCCGACCAAAACAGCGCAGC-----TTATGAA 172  
Db 126 GCTGATCATGAGCGGTCAACAACCTGCGTGGCGCGTTTGGCCCAATAACCGCAGCGTGTCTGGA 185  
QY 173 TGAATTTACGCCAGATTAAGTGGCCACATGCGCGGTAAAGCGGGGTATTTAAGCCACG 232  
Db 186 ACATCCACCGCAAGCGCTTGGCGGAGTGGAGTGTCCCGGGCGGCGCAGCTGACCAACAA 245  
QY 233 TGGTGGGAGTGTGTCATCATTTGATGGGGGGT- TTTATGTTGATTTACTTTTCGAGCCTTG 291  
Db 246 AGGCGGTGTGCTGGAGGTGTATATGGCCACTATATGCGGCAATGGCTGGCGCAGCAGAA 305  
QY 292 GTTTGTTAGCGGG- CGATGTCGGCAGAGGGGGGTATATGCAAGCAGATATCGA 349  
Db 306 GCTGGTACCAAGCGCGGAGTGGCGCGGAAACGCGGTATTATGCTTACGCTAACACGCT 365  
QY 350 TCAACGTAACCGCTTAAACCGGACAGGCAATTTCTTGTGTTGTTGCTCCGGGGTGTGTTT 409  
Db 366 GCAGCGCACCGTCCGACCGCGCAGTTCTTTATCACCGCGCGTTCGCGGCTGGGCTAT 425  
QY 410 GACCGTGCATATCAGCGCGGATTTGAAAAAGACGATCCCTGTTCCATCCGTTAGAGAC 469  
Db 426 CCGGTGCATACACAGCTCAGATGGGCACCATGGACCCGACCTTCAACCGGTTATTAC 485  
QY 470 TGGCGTGTGAAGTTAGACAAACGCAACAGATAAAGCGATTGAAGAA 518  
Db 486 CGACGACTCCCGCGCTTTCGCGAGAAAGCGCTGCAGCGCATGGAATAA 534

RESULT 7  
US-09-489-039A-6330  
; Sequence 6330, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489, 039A  
; PRIOR FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 6330  
; LENGTH: 1569  
; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-6330

Query Match 3.6%; Score 48; DB 4; Length 1569;  
Best Local Similarity 48.0%; Pred. No. 1.4e-05;  
Matches 170; Conservative 0; Mismatches 180; Indels 4; Gaps 1;

QY 73 GCGCGGTAGCGCGGAGCCATCGGCTATACCTTTAGAACGTGTGTTATTTTGGTTCG 132  
Db 364 GCGTGCATCTGCGCGCGCGGAGCTGGACCTGGAGAAAGTGTGAGCTCAGCGC 423  
QY 133 CATGCTTTCGCTGCCGA-----CCAAACACGAGCTTATGAATGATGTTAGCCAGAT 188  
Db 424 CACGTTATTCGTCGCGAGCGCGGCAACCGGAGCCATCGAGCGCGCCACCGCGCGA 483  
QY 189 AAGTGGCCACAATGCGCGGTAAAGCGGGTATTTAACGCCAGCTGTGGGAGTTGGTC 248  
Db 484 CCGTGGACCGAGTGGACCCCATGACGGGAGCTCACCGGCATGCTATGCGCGCTG 543  
QY 249 ACATTGATGGGGGGTTTTATGTTGATTTCTTCGAGCGCTTGTGTTTGTAGCGCGGA 308  
Db 544 GTCAACAAGGGCGTGGGAAAGCGCAGCAATTACCGCGAGCTCGGCTGTGCGAGCGGA 603  
QY 309 TGTCCGCGAGAGGGGGGTATATGACAGGAGATATCGATCAACGATCCCGTTAAC 368  
Db 604 TGCCCGAGCGGAGTGCATATATACGTGCGCGCCAGCCCCGCTGCAGCGGACCGAGGACC 663

QY 369 GGACAGCATTTCTTGTATGGTGTGGCTCCGGGGTGTGGTTTGACCGTGCATAAT 422

D<sub>b</sub> 664 GCCCAGCGCGCTGTGGATGGCGCCTTCCCGGCTCGGGCGGTATCCATTAT 717

## RESULT 8.

```

US-08-920-812-23
; Sequence 23, Application US/08920812
; Patent No. 5763186
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotosugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,812
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5975 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORGANISM: Klebsiella pneumoniae
; STRAIN: Clinical Isolate KI-50
US-08-920-812-23

```

Query Match	3.6%;	Score 48;	DB 1;	Length 5975;
Best Local Similarity	48.0%;	Pred. No. 3.3e-05;		
Matches 170;	Conservative 0;	Mismatches 180;	Indels 4;	Gaps 1;
QY	73	GCGCCGTAGCCGCCGACCATCGGGCTATATCTTTAGAACGTGTGGTATTATTTTCAGTCGC	132	
Db	623	CGCGTGCATCTGCCGCCGCCGCGACCTGGCAGCTGGAGAAAGTGTGAGCTCAGCCGC	682	
QY	133	CATGGTCTCGCTCCCGA-----CCAAACAACGCAGCTTATGAATGATGTTTACGCCAGAT	188	
Db	683	CACGGTATTCGTCCGCCGACGGCCGGCAACCGGGAAGCATCGAGGCGCCACCGGCCGA	742	
QY	189	AAGTGGCCACAATGGCCGGTAAAGCGGGTATTTAACGCCACCTGGTGGGAGTTGGTC	248	
Db	743	CCGTGGACCGAGTGGACCAACCCATGAACGGGAGCTCACCGGCCCATGGCTATGCGCCCGTG	802	
QY	249	ACATTGATGGGGGGTTTATGTGCAATTACTTTCGCAGCCCTTGGTTTGTATTAGCCGCCGGA	308	

803	GTCAACAAGGCGTGGCGAAGGCCACGATTA	CGGCAGCTCGCCTGCTGCAGGCGCGA	862
Db			
309	TGTCGCGCAGAGGGGGGGTATATGCA	CAGCGAGATATCGATCAACGTACCCGCTTAACC	368
Qy			
863	TCGCCGACGGCGGAGTCGATATAGTGGG	CCAGCCCGCTGCAGCGACGCGAGCGACC	922
Db			
369	GGACAGGCATTCTTTGATGTGTGGCTCCG	GGGTGTGTTTGACCGTGCATAAT	422
Qy			
923	GCCACAGGCGCTGTTGATGCGCGCTTC	CCCGGCTGCGGCTGCCTATCAATAT	976
Db			

## RESULT 9

US-08-920-827-23  
; Sequence 23, Application US/08920827  
; Patent No. 5770375  
; GENERAL INFORMATION:  
; APPLICANT: Ohno, Tsuneya  
; APPLICANT: Matsuhisa, Akio  
; APPLICANT: Uehara, Hirotosugu  
; APPLICANT: Eda, Soji  
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/920,827  
; FILING DATE: 29-AUG-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,577  
; FILING DATE: 27-MAR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rin-Laures, Li-Hsien  
; REGISTRATION NUMBER: 33,547  
; REFERENCE/DOCKET NUMBER: 19036/32420  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5975 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
; ORIGINAL SOURCE:  
; ORGANISM: Klebsiella pneumoniae  
; STRAIN: Clinical Isolate KI-50  
; US-08-920-827-23

	Query Match	3.6%	Score 48	DB 1	Length 5975	
	Best Local Similarity	48.0%	Pred. No. 3.3e-05			
	Matches 170	Conservative 0	Mismatches 180	Indels 4	Gaps 17	
QY	73	GCGCCGTTAGCCGCGAGCCATCGGGCTATCTTTAGAACGTGTGGTTATTTTGAGTCCG	132			
DB	623	GCGETGCAATCTCGCGCGCGCGAGCTGGCAGCTGGAGAAAGTGTGTGAGCTCAGCCGC	682			
QY	133	CATGGTGTTCCTCGCCGA-----CAAACAACGCGAGCTTATGAATGATGTTTACGCCAGAT	188			
DB	683	CACGCTATTCCTCGCGCGAGCGCCGGCAACCGGAAGCCATCGAGGCGCCACCGCGCCGA	742			

QY 189 AAGTGGCCCAATGCGCGGTAAAGCGGGTATTTAAACCCACGCTGGTGGAGTTGGTC 248  
 Db 743 CGGTGACCGAGTGGACCAACCCATGACGGGAGCTCACCGGCATGGCTATGCCCGCGTG 802  
 QY 249 ACATTGATGGGGGGTTTATGTTGTTACTTTCGAGCGCTTGGTTTGTAGCGGGGA 308  
 Db 803 GTCAACAAGGGGGTCCGGAAGCCAGCATACCGCCAGCTGGCTGCTGAGCGCCGA 862  
 QY 309 TGTCCCGGACAGAGGGGGGTATATGACAGGCGAGATATCGATCAACGTAACCGCTTAACC 368  
 Db 863 TGCCCGACGGGAGTGCATATACGTGCGCGCACCCCTGCAGCGGAGCGGAGCC 922  
 QY 369 GAGACGCGATTTCTTGATGTTGGCTCGGGGTGGTTTACCGTGCATTAAT 422  
 Db 923 GCCACGGCGCTGGTGGATGGCGCTTCCCGCGCTGCGGCTCGCTATCCATTAT 976

RESULT 10

US-08-921-177-23  
 ; Sequence 23, Application US/08921177  
 ; Patent No. 5798211  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ohno, Tsuneya  
 ; APPLICANT: Matsuhisa, Akio  
 ; APPLICANT: Uehara, Hirotsugu  
 ; APPLICANT: Eda, Soji  
 ; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
 ; NUMBER OF SEQUENCES: 25  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
 ; STREET: 6300 Sears Tower, 233 South Wacker Drive  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: United States of America  
 ; ZIP: 60606-6402  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/921,177  
 ; FILING DATE: 29-AUG-1997  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/362,577  
 ; FILING DATE: 27-MAR-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Rin-Laures, Li-Hsien  
 ; REGISTRATION NUMBER: 33,547  
 ; REFERENCE/DOCKET NUMBER: 19036/32420  
 ; TELEPHONE: 312/474-6300  
 ; TELEFAX: 312/474-0448  
 ; TELEX: 25-3856  
 ; INFORMATION FOR SEQ ID NO: 23:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 5975 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: Genomic DNA  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Klebsiella pneumoniae  
 ; STRAIN: Clinical Isolate KI-50  
 ; US-08-921-177-23

Query Match 3.6%; Score 48; DB 1; Length 5975;  
 Best Local Similarity 48.0%; Pred. No. 3.3e-05;  
 Matches 170; Conservative 0; Mismatches 180; Indels 4; Gaps 1;  
 QY 73 GCGCGGTAGCGCGGACCATCGGCTATATTTAGACAGTGGTTATTTAGTCGC 132

Db 623 GCGTGAATCTGCGCGCGCGGACTGGAGCTGGAGAAAGTGTGAGCTCAGCCGC 682  
 QY 133 CATGCTGTGCTCGCCGA-----CCAAACAACGAGCTTATGATGATGTTAGCCAGAT 188  
 Db 683 CACGTTATTCGTCGCGGACGCGCGCAACCGGAGCCATCGAGGCGCGCACCGGCGGA 742  
 QY 189 AAGTGGCCCAATGCGCGGTAAAGCGGGTATTTAAACCCACGCTGGTGGAGTTGGTC 248  
 Db 743 CCCTGGAACCGAGTGGACCCATGACGGGAGCTCACCGGCCATGGCTATGCCCGCGTG 802  
 QY 249 ACATTGATGGGGGGTTTATGTTGTTACTTTCGAGCGCTTGGTTTGTAGCGCGGGA 308  
 Db 803 GTCAACAAGGGGGTCCGGAAGCCAGCATACCGCCAGCTGGCTGCTGAGCGCCGA 862  
 QY 309 TGTCCCGGACAGAGGGGGGTATATGACAGGCGAGATATCGATCAACGTAACCGCTTAACC 368  
 Db 863 TGCCCGACGGGAGTGCATATACGTGCGCGCACCCCTGCAGCGGAGCGGAGCC 922  
 QY 369 GAGACGCGATTTCTTGATGTTGGCTCGGGGTGGTTTACCGTGCATTAAT 422  
 Db 923 GCCACGGCGCTGGTGGATGGCGCTTCCCGCGCTGCGGCTCGCTATCCATTAT 976

RESULT 11

US-08-362-577C-23  
 ; Sequence 23, Application US/08362577C  
 ; Patent No. 5807673  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ohno, Tsuneya  
 ; APPLICANT: Matsuhisa, Akio  
 ; APPLICANT: Uehara, Hirotsugu  
 ; APPLICANT: Eda, Soji  
 ; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
 ; NUMBER OF SEQUENCES: 25  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
 ; STREET: 6300 Sears Tower, 233 South Wacker Drive  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: United States of America  
 ; ZIP: 60606-6402  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/362,577C  
 ; FILING DATE: 27-MAR-1995  
 ; CLASSIFICATION: 536  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Rin-Laures, Li-Hsien  
 ; REGISTRATION NUMBER: 33,547  
 ; REFERENCE/DOCKET NUMBER: 19036/32420  
 ; TELEPHONE: 312/474-6300  
 ; TELEFAX: 312/474-0448  
 ; TELEX: 25-3856  
 ; INFORMATION FOR SEQ ID NO: 23:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 5975 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: Genomic DNA  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Klebsiella pneumoniae  
 ; STRAIN: Clinical Isolate KI-50  
 ; US-08-362-577C-23

Query Match 3.6%; Score 48; DB 1; Length 5975;  
 Best Local Similarity 48.0%; Pred. No. 3.3e-05;

Matches 170; Conservative 0; Mismatches 180; Indels 4; Gaps 1;  
QY 73 GCGCGGTAGCGCGGAGCCATCGGCTATCTTTAGAACGTGTGTTATTTTGAAGTCG 132  
Db 623 GCGGTGCAATCTGCGCGCGCGGAGCTGGCAGCTGGAGAAAGTGTGAGCTCAGCCGC 682  
QY 133 CATGTGTTCCTCGCCGA-----CCAAACAACGAGCTTATGATGATTTAGCCAGAT 188  
Db 693 CACGTATTCGTCCCGGAGCGCGGAGCCGAGCAACCGGAGGCAATCGAGGCGGCAACCGCGCGA 742  
QY 189 AAGTGGCCCAATCGCGGTAAAGCGGGTATTTAAAGCCACGCTGTGTGGGAGTTGGTC 248  
Db 743 CCGTGGACCGAGTGGACCAACCATGACGCGGAGCTCACCGGCCATGCGTATGCGCGCGTG 802  
QY 249 ACATTGATGCGGGGTTTATGCTGATTTACGAGCTTCTGCGAGCTTGTGTTAGCGCGCGGA 308  
Db 803 GTCACAAAGGCGGTGCGGAGGCGGAGCAATACCGCCAGCTCGCGCTGCTGCGAGCGCGGA 862  
QY 309 TGTCCGGCAGAGGGGGGTATATGACAGGAGATATCGATCAACGTACCGGCTTAACC 368  
Db 863 TGCCCGACGGCGAGTGATATACGTGCGCGCCAGCCGCTGACGAGCGGAGCGGAGCC 922  
QY 369 GGACAGGCAATCTTGTGATGCTGCGGCTCGGGGTGTGTTTGAACGCTGCAAT 422  
Db 923 GCCCAGGCGCTGCTGATGCGGCTTCCCGGCTGCGGCTGCTATCCATTAT 976

## RESULT 12

US-08-920-828-23  
; Sequence 23, Application US/08920828

; Patent No. 5853598

; GENERAL INFORMATION:

; APPLICANT: Ohno, Teuneya

; APPLICANT: Matsuhisa, Akio

; APPLICANT: Uehara, Hirotsugu

; APPLICANT: Ega, Soji

; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease

; NUMBER OF SEQUENCES: 25

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

; STREET: 6300 Sears Tower, 233 South Wacker Drive

; CITY: Chicago

; STATE: Illinois

; COUNTRY: United States of America

; ZIP: 60606-6402

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/920,828

; FILING DATE: 29-AUG-1997

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/362,577

; FILING DATE: 27-MAR-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Rin-Laures, Li-Hsien

; REGISTRATION NUMBER: 33,547

; REFERENCE/DOCKET NUMBER: 19036/32420

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 312/474-6300

; TELEFAX: 312/474-0448

; TELEX: 25-3856

; INFORMATION FOR SEQ ID NO: 23:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 5975 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: Genomic DNA

; ORIGINAL SOURCE:

; ORGANISM: Klebsiella pneumoniae  
; STRAIN: Clinical Isolate KI-50  
US-08-920-828-23

Query Match 3.6%; Score 48; DB 2; Length 5975;  
Best Local Similarity 48.0%; Pred. No. 3.3e-05;

Matches 170; Conservative 0; Mismatches 180; Indels 4; Gaps 1;

QY 73 GCGCGGTAGCGCGGAGCCATCGGCTATCTTTAGAACGTGTGTTATTTTGAAGTCG 132  
Db 623 GCGGTGCAATCTGCGCGCGCGGAGCTGGCAGCTGGAGAAAGTGTGAGCTCAGCCGC 682  
QY 133 CATGTGTTCCTCGCCGA-----CCAAACAACGAGCTTATGATGATTTAGCCAGAT 188  
Db 693 CACGTATTCGTCCCGGAGCGCGGAGCCGAGCAACCGGAGGCAATCGAGGCGGCAACCGCGCGA 742  
QY 189 AAGTGGCCCAATCGCGGTAAAGCGGGTATTTAAAGCCACGCTGTGTGGGAGTTGGTC 248  
Db 743 CCGTGGACCGAGTGGACCAACCATGACGCGGAGCTCACCGGCCATGCGTATGCGCGCGTG 802  
QY 249 ACATTGATGCGGGGTTTATGCTGATTTACGAGCTTCTGCGAGCTTGTGTTAGCGCGCGGA 308  
Db 803 GTCACAAAGGCGGTGCGGAGGCGGAGCAATACCGCCAGCTCGCGCTGCTGCGAGCGCGGA 862  
QY 309 TGTCCGGCAGAGGGGGGTATATGACAGGAGATATCGATCAACGTACCGGCTTAACC 368  
Db 863 TGCCCGACGGCGAGTGCATATACGTGCGCGCCAGCCGCTGACGAGCGGAGCGGAGCC 922  
QY 369 GGACAGGCAATCTTGTGATGCTGCGGCTCGGGGTGTGTTTGAACGCTGCAAT 422  
Db 923 GCCCAGGCGCTGCTGATGCGGCTTCCCGGCTGCGGCTGCTATCCATTAT 976

## RESULT 13

US-08-232-463-14/c

; Sequence 14, Application US/08232463

; Patent No. 5670367

; GENERAL INFORMATION:

; APPLICANT: DORNER, F.

; APPLICANT: SCHEIFLINGER, F.

; APPLICANT: FALKNER, F. G.

; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS

; NUMBER OF SEQUENCES: 52

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: 1800 Diagonal Road, Suite 500

; CITY: Alexandria

; STATE: VA

; COUNTRY: USA

; ZIP: 22313-0299

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/232,463

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/07/935,313

; FILING DATE:

; APPLICATION NUMBER: EP 91 114 300.6

; FILING DATE: 26-AUG-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: BENT, Stephen A.

; REGISTRATION NUMBER: 29,768

; REFERENCE/DOCKET NUMBER: 30472/114 IMMU

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703)836-9300

; TELEFAX: (703)883-4109

; TELEX: 899149

; INFORMATION FOR SEQ ID NO: 14:

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PT9pt-F18
US-08-232-463-14

Query Match      2.8%; Score 37; DB 1; Length 7218;
Best Local Similarity 2.6%; Pred. NO. 0.2;
Matches 10; Conservative 208; Mismatches 163; Indels 0; Gaps 0;

QY 176 TGTACGCCAGTAAGTGCACATGCGCGTAAAGCGGGGTATTTAAACGCCACGTGG 235
DB 1440 TGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1381
QY 236 TCGGAGTGTGTCACATTCATCGGGGGTTCATCGTACTTTCACGCTTGGTTT 295
DB 1380 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1321
QY 296 GTTACGGCGGATGTCGGCAGAGGGGGGTATATGACAGGCGAGATATCGATCAAC 355
DB 1320 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1261
QY 356 TACCGCTTAACCGACAGCATTTCTTGATGCTGCTCGGGGTGTTTGACCGT 415
DB 1260 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1201
QY 416 GCATATATCAGCGGATTTGAAAAGACCGATCCCTGTTCATCGGTAGAGACTGGGT 475
DB 1200 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1141
QY 476 GTGTAAGTTAGACACGACACAGATAAAGCGATTGAGAACGATTGGCGGGCGCTT 535
DB 1140 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1081
QY 536 AGATACGCTGACCGCGCTA 556
DB 1080 RRRRRRRRRRRATCGCAA 1060

RESULT 14
US-09-252-991A-4062/c
; Sequence 4062, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4062
; LENGTH: 1602
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4062

Query Match      2.6%; Score 35; DB 4; Length 1602;
Best Local Similarity 55.3%; Pred. NO. 0.36;
Matches 68; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 295 TGTACGGCGGGATGTCGGCAGAGGGGGGTATATGACAGGCGAGATATCGATCAAC 354
DB 539 TGTTCGGCGGGAACCGCGAGGCGATCGGTCTAGCGGACAGGTCTGAGTCTGATCAGC 480
QY 355 GTACCGCTTAACCGACAGGCATTTCTTGATGCTGCTCGGGGTGTTTGACCG 414
```

```
DB 479 GTGCGGAGCCCTCGGGATGGCGGCTCTGCTCTGCGCGGACGGGAGGTCTGCACC 420
QY 415 TGC 417
DB 419 AGC 417

RESULT 15
US-09-252-991A-4093
; Sequence 4093, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4093
; LENGTH: 1773
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4093

Query Match      2.6%; Score 35; DB 4; Length 1773;
Best Local Similarity 55.3%; Pred. NO. 0.39;
Matches 68; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 295 TGTACGGCGGGATGTCGGCAGAGGGGGGTATATGACAGGCGAGATATCGATCAAC 354
DB 1094 TGTTCGGCGGGAACCGCGAGGCGATCGGTCTAGCGGACAGGTCTGAGTCTGATCAGC 1153
QY 355 GTACCGCTTAACCGACAGGCATTTCTTGATGCTGCTCGGGGTGTTTGACCG 414
DB 1154 GTGCGGAGCCCTCGGGATGGCGGCTCTGCTCTGCGCGGACGGGAGGTCTGCACC 1213
QY 415 TGC 417
DB 1214 AGC 1216

Search completed: May 3, 2004, 03:58:21
Job time : 76.6484 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 3, 2004, 01:47:38 ; Search time 568.19 Seconds  
(without alignments)  
10548.559 Million cell updates/sec

Title: US-10-021-723B-3  
Perfect score: 1325  
Sequence: 1 atgtcagattagaaatcg.....aacctgcctgcacatctaa 1325

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2936184 seqs, 2261732022 residues

Total number of hits satisfying chosen parameters: 5872368

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/FCI\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1325	100.0	1325	15	US-10-021-723A-3
2	1306	98.6	1326	13	US-10-021-723A-3
3	1292.2	97.5	1325	15	US-10-021-723A-1
4	827.6	62.5	1326	15	US-10-021-723A-11
5	242	18.3	1281	15	US-10-334-672-4
6	242	18.3	1281	15	US-10-334-671-4
7	226.2	17.1	1308	15	US-10-156-660-1
8	225.2	17.0	1301	9	US-09-866-379-5
9	225.2	17.0	1301	9	US-09-866-379-6
10	225.2	17.0	1301	9	US-09-866-379-7
11	225.2	17.0	1301	15	US-10-156-660-3
12	224.6	17.0	1299	13	US-10-282-122A-7167
13	223.6	16.9	1301	9	US-09-866-379-9
14	218.8	16.5	1323	9	US-09-777-566A-1

## ALIGNMENTS

RESULT 1  
US-10-021-723A-3  
; Sequence 3, Application US/10021723A  
; Publication No. US20030101476A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay  
; APPLICANT: Mathur, Eric  
; APPLICANT: Richardson, Toby  
; APPLICANT: Robertson, Dan  
; APPLICANT: Barton, Nelson  
; TITLE OF INVENTION: Recombinant Phytases and Uses Thereof  
; FILE REFERENCE: 112766.140 (DIV-016CIP)  
; CURRENT APPLICATION NUMBER: US/10/021,723A  
; CURRENT FILING DATE: 2002-10-23  
; PRIOR APPLICATION NUMBER: US 60/255,090  
; PRIOR FILING DATE: 2000-12-12  
; NUMBER OF SEQ ID NOS: 74  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 1325  
; TYPE: DNA  
; ORGANISM: Yersinia pestis  
US-10-021-723A-3

Query Match 100.0%; Score 1325; DB 15; Length 1325;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGTCAGTATTAGAAATCGTACGGCTATCTCGGCTGTTATGATGCTTAAGCGGATTG 60  
Db 1 ATGTCAGTATTAGAAATCGTACGGCTATCTCGGCTGTTATGATGCTTAAGCGGATTG 60  
Qy 61 GCTGCTATTACTCGCGGTAGCCGCGGAGCCATCGGCTTACTTTAGAACGTTGGTT 120  
Db 61 GCTGCTATTACTCGCGGTAGCCGCGGAGCCATCGGCTTACTTTAGAACGTTGGTT 120  
Qy 121 ATTTTGATCGCCATGTTGTTCTGTCGCCGACCAACACGAGCTTATGATGATGTTA 180

```

Db 121 ATTTGAGTCGCCATGTTGTCGTCGCGCAACAAACGCGAGCTTATGAATGATGTTA 180
QY 181 CGCAGATTAAGTGGCCCAATAGGCGCGTAAAGCGGGTATTTAAAGCCACGCTGGTGGCG 240
Db 181 CGCAGATTAAGTGGCCCAATAGGCGCGTAAAGCGGGTATTTAAAGCCACGCTGGTGGCG 240
QY 241 AGTTGGTCAATGATGGGGGGGTTTATAGTGATTAATCTTTCGAGCGCTTGGTTGTTAG 300
Db 241 AGTTGGTCAATGATGGGGGGGTTTATAGTGATTAATCTTTCGAGCGCTTGGTTGTTAG 300
QY 301 CGCGGGATGTCGCGCAGAGCGGGGTATATGACAGGAGATATCGATCAACGATACCC 360
Db 301 CGCGGGATGTCGCGCAGAGCGGGGTATATGACAGGAGATATCGATCAACGATACCC 360
QY 361 GCTTAACCGGACAGGCAATCTTCTGATGTTGGTTCGCGGGTGGTGGTGGTGGTGGTGGT 420
Db 361 GCTTAACCGGACAGGCAATCTTCTGATGTTGGTTCGCGGGTGGTGGTGGTGGTGGTGGT 420
QY 421 ATCAGGCGGATTTGAAAGAACCGATCCCTGTTCCATCCGTTAGAGACTGGCGTGTTA 480
Db 421 ATCAGGCGGATTTGAAAGAACCGATCCCTGTTCCATCCGTTAGAGACTGGCGTGTTA 480
QY 481 AGTTAGACAACGCAAAACAGATTAAGGATTAAGAACGATTTGGGCGGCGCGTTAGATA 540
Db 481 AGTTAGACAACGCAAAACAGATTAAGGATTAAGAACGATTTGGGCGGCGCGTTAGATA 540
QY 541 CGGTGAGCGAGCGCTAGCTAAACCTTTTCCAGATGGGGAGCGTCTGAATTTGCGG 600
Db 541 CGGTGAGCGAGCGCTAGCTAAACCTTTTCCAGATGGGGAGCGTCTGAATTTGCGG 600
QY 601 CTTCTCCTTATGCAATCTTTGCAACAGCAAGGAAAGAAAGCGTGTGATTTGCCACCTTTG 660
Db 601 CTTCTCCTTATGCAATCTTTGCAACAGCAAGGAAAGAAAGCGTGTGATTTGCCACCTTTG 660
QY 661 CGGCAATGAATTAAGGCGGCGGAGATTTGGTATCTTTTATGATTTGCGGCGCTAGG 720
Db 661 CGGCAATGAATTAAGGCGGCGGAGATTTGGTATCTTTTATGATTTGCGGCGCTAGG 720
QY 721 CGTTATCATCGACATGCGGTGAATCTTTTATTAACAAACGCAAGGAAAGCGTGTGATTTGCCAGG 780
Db 721 CGTTATCATCGACATGCGGTGAATCTTTTATTAACAAACGCAAGGAAAGCGTGTGATTTGCCAGG 780
QY 781 TTGCTGCGCAACGCTTAAGGCGGCGGAGATTTGGTATCTTTTATGATTTGCGGCGCTAGG 840
Db 781 TTGCTGCGCAACGCTTAAGGCGGCGGAGATTTGGTATCTTTTATGATTTGCGGCGCTAGG 840
QY 841 CGCAATTTAATTAAGGCAAAACACCGTATATCGCCCGTCAATAAGGAGCGCATTTAT 900
Db 841 CGCAATTTAATTAAGGCAAAACACCGTATATCGCCCGTCAATAAGGAGCGCATTTAT 900
QY 901 TACAGCAGATAGATACGCGCTTTAACCTTCACTGATGCTCAGGCGGCAAAAGCTACCCA 960
Db 901 TACAGCAGATAGATACGCGCTTTAACCTTCACTGATGCTCAGGCGGCAAAAGCTACCCA 960
QY 961 TTTGAGCCCAAAACCGGCTCTGTTCTCGTGGGCGATGATACCAATATTGCCAATATTG 1020
Db 961 TTTGAGCCCAAAACCGGCTCTGTTCTCGTGGGCGATGATACCAATATTGCCAATATTG 1020
QY 1021 CGGATGCTGGGAGCGGATGCGAGCTACCCGAGCAACCTGATTAATCTCCACCGAGTG 1080
Db 1021 CGGATGCTGGGAGCGGATGCGAGCTACCCGAGCAACCTGATTAATCTCCACCGAGTG 1080
QY 1081 GGGGATTTGTTTGAACCTATGGCAAAATCCGATTAACCCAGCGCTAGTTGGGTGA 1140
Db 1081 GGGGATTTGTTTGAACCTATGGCAAAATCCGATTAACCCAGCGCTAGTTGGGTGA 1140
QY 1141 AGATGTTCTACCAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
Db 1141 AGATGTTCTACCAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
QY 1201 ACCAGCGGATTAATTTCCGTTGCGATGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1260

```

```

Db 1201 ACCAGCGGATTAATTTCCGTTGCGATGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1260
QY 1261 TTTTGGAGCTTGATACATTTCCAAAGAAAGTGGCTTAAGGTAATTAAGAACTGCTGCCACA 1320
Db 1261 TTTTGGAGCTTGATACATTTCCAAAGAAAGTGGCTTAAGGTAATTAAGAACTGCTGCCACA 1320
QY 1321 TCTAA 1325
Db 1321 TCTAA 1325

RESULT 2
US-10-282-122A-41608
; Sequence 41608, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 41608
; LENGTH: 1326
; TYPE: DNA
; ORGANISM: Yersinia pestis
US-10-282-122A-41608

Query Match 98.6%; Score 1306; DB 13; Length 1326;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1320; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 1 ATGTGATATTAGAAAATCGTGTACGGCTATCTGGGCTGTTATTCATGCTAAGCGGATTG 60
Db 1 ATGTGATATTAGAAAATCGTGTACGGCTATCTGGGCTGTTATTCATGCTAAGCGGATTG 60
QY 61 GCTGCTATTACTGCGCGGCTAGCCCGGAGCCATCGGCTATCTTTTAGAAGCTGTGTT 120
Db 61 GCTGCTATTACTGCGCGGCTAGCCCGGAGCCATCGGCTATCTTTTAGAAGCTGTGTT 120

```



121 ATTTGAGTCGCCATGGTGTTCCTCGCCGACCAAC-AACGAGCTTATGATGATTT 179  
121 ATTTGAGTCGCCATGGTGTTCCTCGCCGACCAACAAACGAGCTTATGATGATTT 180  
180 ACGCCAGATGAAGTGGCCCAATGCGCGGTAAAGCGGGGTATTTAAAGCGCAAGTGGTGG 239  
181 ACGCCAGATGAAGTGGCCCAATGCGCGGTAAAGCGGGGTATTTAAAGCGCAAGTGGTGG 240  
240 GAGTGGTCACATGATGGGGGGTTTTATGGTGATATCTTTCGACGCTTGGTTTGTTA 299  
241 GAGTGGTCACATGATGGGGGGTTTTATGGTGATATCTTTCGACGCTTGGTTTGTTA 300  
300 GCGGCGGGATGTCGCGCAGAGGGGGGTATATGCAAGCGAGATATCGATCAACGATCC 359  
301 GCGGCGGGATGTCGCGCAGAGGGGGGTATATGCAAGCGAGATATCGATCAACGATCC 360  
360 CGCTTAACCGGACAGGCAATTTCTTGATGCTGGCTCCGGGGTGTGTTTGAACGCTGAT 419  
361 CGCTTAACCGGACAGGCAATTTCTTGATGCTGGCTCCGGGGTGTGTTTGAACGCTGAT 420  
420 AATCAGCGCCGATTTGAAAGAGCGATCCCTCTGTCATCCGTAGAGACTGGCGTGTCT 479  
421 AATCAGCGCCGATTTGAAAGAGCGATCCCTCTGTCATCCGTAGAGACTGGCGTGTCT 480  
480 AAGTTAGACAAAGCAAGATGAAGCAATTTGAAGCAATTTGGGCGGGCGGCTTAGAT 539  
481 AAGTTAGACAAAGCAAGATGAAGCAATTTGAAGCAATTTGGGCGGGCGGCTTAGAT 540  
540 ACGGTAGCAGCGCTACGCTAAACCTTTTGGCCAGATGGGAGCGTGTGATTTTGG 599  
541 ACGGTAGCAGCGCTACGCTAAACCTTTTGGCCAGATGGGAGCGTGTGATTTTGG 600  
600 GCTTCTCTTATTTGCAATCTTTGCAACAGCAAGGAAACAGTGTGATTTTGGCCACTTT 659  
601 GCTTCTCTTATTTGCAATCTTTGCAACAGCAAGGAAACAGTGTGATTTTGGCCACTTT 660  
660 GCGGCCAATGAGTTAAGCTTAATGAAGAGGACCAAGTGAACCTCTGAGTGGGCGCACATG 719  
661 GCGGCCAATGAGTTAAGCTTAATGAAGAGGACCAAGTGAACCTCTGAGTGGGCGCACATG 720  
720 GCGTTATCATCGATGGGTGAAATCTTCTTATTAACAAACGCAAGCCATGCGCAGAG 779  
721 GCGTTATCATCGATGGGTGAAATCTTCTTATTAACAAACGCAAGCCATGCGCAGAG 780  
780 GTTGCCTGGCAACGGCTAAAGGGGGGAGAAATGGGTATCCTTATTTGATTAACATAAC 839  
781 GTTGCCTGGCAACGGCTAAAGGGGGGAGAAATGGGTATCCTTATTTGATTAACATAAC 840  
840 GCGCAATTTAATTTAATGCAAAACACCGCTATATCGCCGTCATTAAGGGAGCGCATTA 899  
841 GCGCAATTTAATTTAATGCAAAACACCGCTATATCGCCGTCATTAAGGGAGCGCATTA 900  
900 TTAAGCAGATAGATAGCGCTTTAAACCTTTCAATCGAGTCTCAGGGGCAAAAGCTACCC 959  
901 TTAAGCAGATAGATAGCGCTTTAAACCTTTCAATCGAGTCTCAGGGGCAAAAGCTACCC 960  
960 ATTTACGCCAAACCGGGTCTTGTCTCGGTGGGATGATACCAATATGCGCAATATT 1019  
961 ATTTACGCCAAACCGGGTCTTGTCTCGGTGGGATGATACCAATATGCGCAATATT 1020  
1020 GCGGGATGCTGGAGCGGACTGGAGTACCGGAGCAACCTGATTAATCTCACCGAGT 1079  
1021 GCGGGATGCTGGAGCGGACTGGAGTACCGGAGCAACCTGATTAATCTCACCGAGT 1080  
1080 GCGGGATGCTGGAGCGGACTGGAGTACCGGAGCAACCTGATTAATCTCACCGAGT 1139  
1081 GCGGGATGCTGGAGCGGACTGGAGTACCGGAGCAACCTGATTAATCTCACCGAGT 1140  
1140 AAGATGTTCTACCAACAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1199  
1141 AAGATGTTCTACCAACAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200  
1200 AACCCAGCGGGTATTTTCCGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1259

1201 AACCCAGCGGGTATTTTCCGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260  
1260 CTTTTCGAGCTTGATACATTTCCAAAGAAAGTGGCTTAAGGTAAATTGAACCTGCTGCCAC 1319  
1261 CTTTTCGAGCTTGATACATTTCCAAAGAAAGTGGCTTAAGGTAAATTGAACCTGCTGCCAC 1320  
1320 ATCTAA 1325  
1321 ATCTAA 1326  
  
RESULT 3  
US-10-021-723A-1  
; Sequence 1, Application US/10021723A  
; Publication No. US20030101476A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay  
; APPLICANT: Mathur, Eric  
; APPLICANT: Richardson, Toby  
; APPLICANT: Robertson, Dan  
; APPLICANT: Barton, Nelson  
; TITLE OF INVENTION: Recombinant Phytases and Uses Thereof  
; FILE REFERENCE: 112766.140 (DIV-016CIP)  
; CURRENT APPLICATION NUMBER: US/10/021.723A  
; CURRENT FILING DATE: 2002-10-23  
; PRIOR APPLICATION NUMBER: US 60/255,090  
; PRIOR FILING DATE: 2000-12-12  
; NUMBER OF SEQ ID NOS: 74  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 1325  
; TYPE: DNA  
; ORGANISM: Yersinia pestis  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 470..487, 491, 492, 522  
; OTHER INFORMATION: n = A,T,C or G  
US-10-021-723A-1  
  
Query Match 97.5%; Score 1292.2; DB 15; Length 1325;  
Best Local Similarity 99.2%; Pred. No. 0;  
Matches 1316; Conservative 0; Mismatches 8; Indels 2; Gaps 2;  
  
QY 1 ATGTCAGTATTAGAAAATCGTGTACGGCTATCTGGGCTGGTATTGATGCTTAAGCGGATTG 60  
DB 1 ATGTCAGTATTAGAAAATCGTGTACGGCTATCTGGGCTGGTATTGATGCTTAAGCGGATTG 60  
QY 61 GCTGCTATTACTGCGCGGTAGCGCCGAGCCATCGGCTATCTTTAGAACGTTGTGTT 120  
DB 61 GCTGCTATTACTGCGCGGTAGCGCCGAGCCATCGGCTATCTTTAGAACGTTGTGTT 120  
QY 121 ATTTTGAAGTCGCCATGGTGTTCGCTCGCCGACCAAC-AACGAGCTTATGAATGATGTT 179  
DB 121 ATTTTGAAGTCGCCATGGTGTTCGCTCGCCGACCAACCAACGAGCTTATGAATGATGTT 180  
QY 180 ACGCCAGATAAGTGGCCCAATGCGCGTAAAGCGGGTATTTAAAGCGCAAGTGGTGG 239  
DB 181 ACGCCAGATAAGTGGCCCAATGCGCGTAAAGCGGGTATTTAAAGCGCAAGTGGTGG 240  
QY 240 GAGTGGTCACATGATGGGGGGTTTTATGGTGATATCTTTCGACGCTTGGTTTGTTA 299  
DB 241 GAGTGGTCACATGATGGGGGGTTTTATGGTGATATCTTTCGACGCTTGGTTTGTTA 300  
QY 300 GCGGCGGGATGTCGCGCAGAGGGGGGTATATGCAAGCGAGATATCGATCAACGATCC 359  
DB 301 GCGGCGGGATGTCGCGCAGAGGGGGGTATATGCAAGCGAGATATCGATCAACGATCC 360  
QY 360 CGCTTAACCGGACAGGCAATTTCTTGATGCTGGCTCCGGGGTGTGTTTGAACGCTGAT 419  
DB 361 CGCTTAACCGGACAGGCAATTTCTTGATGCTGGCTCCGGGGTGTGTTTGAACGCTGAT 420  
QY 420 AATCAGCGCGATTTGAAAGAGCGATCCCTCTGTCATCCGTAGAGACTGGCGTGTCT 479

421	AATCAGCGCGATTGAAAAAGACCAGATCCCCTGTTCCATCCGGTAGAGNTGGCGTGCT	480
480	AAGTTAGACAACGCACAAACAGATAAAGCGATTGAAGAACAGATTGGGCGGGCGGTAGAT	539
481	AAGTTANACGNNGCACAAACAGATAAAGCGATTGAAGAACANTTTGGGCGGGCGGTAGAT	540
540	ACGGTGAACCCAGCGCTACGCTAAACCTTTTGGCCAGATGGGCGACGTCTGAATTTTCGG	599
541	ACGGTGAACCCAGCGCTACGCTAAACCTTTTGGCCAGATGGGCGACGTCTGAATTTTCGG	600
600	GCTTCTCTTATTTGCAAAATCTTTTGCAACAGCAAGGAAAAACGTGTGATTTTGGCCACTTT	659
601	GCTTCTCTTATTTGCAAAATCTTTTGCAACAGCAAGGAAAAACGTGTGATTTTGG-CCACTTT	659
660	CGCGCCAAATGAAGTTAAAGCTTAATAAAGAGGGACAAAGTCACCTCAGTGGGGCACTG	719
660	CGCGCCAAATGAAGTTAAAGCTTAATAAAGAGGGACAAAGTCACCTCAGTGGGGCACTG	719
720	CGGTTATCATCGACATTTGGGTGAATTCCTTTATTAACAAACGCAACAGCCATGCCAGAG	779
720	CGGTTATCATCGACATTTGGGTGAATTCCTTTATTAACAAACGCAACAGCCATGCCAGAG	779
780	GTTGGCTCGCAACGGCTAAAAAGGGGCGGAGAAATCGGTATCCTTAATGTCTATTACATAAC	839
780	GTTGGCTCGCAACGGCTAAAAAGGGGCGGAGAAATCGGTATCCTTAATGTCTATTACATAAC	839
840	CGCGAAATTAATTTAATGGCAAAAAACACGTTATATCGCCCGTCAATAAAGGGACGCCATTA	899
840	CGCGAAATTAATTTAATGGCAAAAAACACGTTATATCGCCCGTCAATAAAGGGACGCCATTA	899
900	TTACAGCAGATAGATACGGCTTTAACCTTTCAACTGGATGCTCAGGGCAGAAAGCTACCC	959
900	TTACAGCAGATAGATACGGCTTTAACCTTTCAACTGGATGCTCAGGGCAGAAAGCTACCC	959
960	ATTTTCAGCCCAAAAACCGGGCTTGTTCTTCGTGGGCATGATACCAATATTGCCAATATT	1019
960	ATTTTCAGCCCAAAAACCGGGCTTGTTCTTCGTGGGCATGATACCAATATTGCCAATATT	1019
1020	CGGGGTATGCTGGGAGCCGACTGGCAGCTACCCGAGCAACCTGATACTCCACCAAGT	1079
1020	CGGGGTATGCTGGGAGCCGACTGGCAGCTACCCGAGCAACCTGATACTCCACCAAGT	1079
1080	GGGGGATTTGGTTTTTGAACATAATGCGAAAAATCCGGATAACCCACCGCTACGTTCCGGTG	1139
1080	GGGGGATTTGGTTTTTGAACATAATGCGAAAAATCCGGATAACCCACCGCTACGTTCCGGTG	1139
1140	AAGATGTTCTACAAACGATGATACGTTGCGTTAATGCGAAAAATTCGATCTGAAAAAT	1199
1140	AAGATGTTCTACAAACGATGATACGTTGCGTTAATGCGAAAAATTCGATCTGAAAAAT	1199
1200	AACCCAGCGGTAATTTATTTCCGTTCCAGTTGCTGTTGTGAATAATACCGTGACGATAAG	1259
1200	AACCCAGCGGTAATTTATTTCCGTTCCAGTTGCTGTTGTGAATAATACCGTGACGATAAG	1259
1260	CTTTTCGAGCTTGATACATTCCAAAGAAAGTGGCTTAGGTTAATTTGAACCTGCCTGCCAC	1319
1260	CTTTTCGAGCTTGATACATTCCAAAGAAAGTGGCTTAGGTTAATTTGAACCTGCCTGCCAC	1319
1320	ATCTAA 1325	
1320	ATCTAA 1325	

RESULT 4  
US-10-021-723A-11  
; Sequence 11, Application US/10021723A  
; Publication No. US20030101476A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay  
; APPLICANT: Mathur, Eric  
; APPLICANT: Richardson, Toby  
; APPLICANT: Robertson, D.

APPLICANT: Barton, Nelson  
 TITLE OF INVENTION: Recombinant Phytases and Uses Thereof  
 FILE REFERENCE: 112766.140 (DIV-016C1P)  
 CURRENT APPLICATION NUMBER: US/10/021,723A  
 CURRENT FILING DATE: 2002-10-23  
 PRIOR APPLICATION NUMBER: US 60/255,090  
 PRIOR FILING DATE: 2000-12-12  
 NUMBER OF SEQ ID NOS: 74  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 11  
 LENGTH: 1326  
 TYPE: DNA  
 ORGANISM: Yersinia pestis  
 S-10-021-723A-11

Query Match	62.5%;	Score	827.6;	DB	15;	Length	1326;
Best Local Similarity	77.0%;	Pred. No.	5.9e-261;				
Matches 1021;	Conservative	0;	Mismatches	304;	Indels	1;	0;
1	ATGTCAGTATTAGAAAATCGTGTACGGCTCTCTGGCTGTTATTTGATGCTTAGCGG						
1	ATGTCGGGACTGGAGAACCGCTCGCCCTTCGGGTTAGTGTAAAGCTGTCCGG						
61	GCTGCTATTACTGCGCCGGTAGCCGCGGAGCCATCGGGCTATACCTTTAGAACGTGTG						
61	GCTGCTATCACCGCGCCTGTGCGCGCGCAACCATCGGGGTACACCTTAGAAACGTGTG						
121	ATTTTGAGTCGCCATGSGTGTTCGCTCCGCGCAACAAACA-ACGCAGGCTTATGATG						
121	ATTTTGAGTCGCCATGSGTGTGCGTAGCCCGACTTAAGCAGACGACGCTGATGAAT						
180	AGCCAGATAAGTGGGCCACATATGGCCGGTAAAGCGGGGTATTTAACGCCACGTGTG						
181	ACACCTGATAAGTGGCCTCAGTGGCCGGTTAAGCGGGCTATTTGACTCCTCGTGG						
240	GAGTTGGTCACATGATGSGGGGGTTTTAAGTGATTACTTTTCGACGCCCTTGSTTT						
241	GAACTGGTCACCTCGATGGCGCGGTTCTATGCGCATTAATTCGCGAGTTTGGGTCT						
300	GCGCGGGGATGTCGGGCAGAGGGGGGTATATGCACAGCGCAGATATCGATCAACCG						
301	GCCGCGGGCTGCCCGCAGAGGGCGGTGTATATGCACAGCGCAGATATCGACGACG						
360	CGGTTAAACCGGACAGGCATTTCTTGATGTGTGGCTCCGGGGTGTGGTTTGACCGG						
361	CGTTTAAACCGGTACGGCTTTCTCGATGGTGTGGCCCGGGTTGCGGCGCTGACTGT						
420	AATCAGGCGGATTTGAAAAAGACCGATCCCTCTGTTCCATCCGGTAGAGACTGGCGG						
421	AATCAGGCGGATCTTAAGAAAACCGATCCTCTGTTTCATCCCGTTGAACCGGGG						
480	AAGTTAGACAAACGCAAAACAGATAAGCGATTGAAGAACGATTTGGCGGGGCCGTT						
481	AAACTGGAACACGCCCAACCCGATAGGCAATTGAGGAAACGCCCTGGCGCGCCGCT						
540	ACGGTGAGCGAGCGCTACGCTAAACCTTTTGCCAGATGGGGACGCTGCTGTAATTT						
541	ACGGTAAAGCAGCGCTATGCCAAACCGTTTTCGCAATGGGCGATGTCTCTGAATCT						
600	GCTTCTCCTTATTTGCAAAATCTTTGCAACAGCAAGGAAAAACGTGTGATTTTGGCC						
601	GCGAGTCCGCTACTGCAAGTCACTGCAAGCAGCAGGGGAAAACTTTGTGACTTCGGAC						
660	GCGGCCAATGAAGTTAAACGTTAATAAGAGGGAGCAAAAAGTAGTACCCTCAGTGGCG						
661	GCGGCCAACGAAGTTAATGTAAACAGGAAGGCAACGAAGTTACCTTGTCAGGCC						
720	GCGTTATCATCGCATTTGGGTGAAATCTTTCTTATTACAAACACGCAACGACCATGCG						
721	GCGCTGTCTAGCACGTTGGCGGCAATCTTCTTGCTGCAGAACCGCAGGCGATGCGC						
780	GTTCCTGCGCAACGGCTAAAGGGCGGAGAAATTTGGGTATCCCTTATTGTCTATTAC						

Db 781 GTAGCGTGGCAGCGTTTGAAGCGCGCTGAGAACTGGGTGTCTCTCTGTGAGCCTGCACAAT 840  
Qy 840 GCGCAATTTAATATGCAAAACACCGTATATCGCCGCTAATAAGGAGCGCATTA 899  
Db 841 GCACAGTTCAACCTGATGGCTAAACGCCATATCATTTGACGCCACAAAAGGCGCGCTT 900  
Qy 900 TTACACAGATAGATACGGCTTTAAACCTTTCAACTGGATGCTTCAGGGGCAAAAGCTACCC 959  
Db 901 TTACAGCAATCGATACCGCACTGACCTTGCACCTGGACGCCCGGCGCAAAAATGCGCG 960  
Qy 960 ATTTACAGCCAAACACCGGGTCTTTGCTCGGTGGGCGATGATACCAATATTTGCCAATATT 1019  
Db 961 ATCTCGGCTCAGAACCGTGTTTTATTTCTCGGGTGGCCAGACACAAAATATTGCTAACATC 1020  
Qy 1020 GCGGGTATGCTGGGAGCGCACTGGCAGCTACCGGAGCACTGATAAATACTCCACAGGT 1079  
Db 1021 GCGGGTATGCTGGGAGCGCACTGGCAGCTACCGGAGCACTGATAAATACTCCACAGGT 1080  
Qy 1080 GGGGGATTGGTTTGAATATGCGCAAAATCCGGATAAACCAACAGCGCTACGTTGGGGTG 1139  
Db 1081 GCGGGTCTGGTCTTTGAGCTGTGGCAGAAATCCGGCAATCATCAACAGTTATGTGGCGGT 1140  
Qy 1140 AGAGTCTTACCAACAGTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1199  
Db 1141 AGAGTCTTATCAGACCAATGATCAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200  
Qy 1200 AACCCAGCGGTATTTTCCGTTGCGAGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1259  
Db 1201 AATCCCGCGCATCATCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260  
Qy 1260 CTTTGGAGCTTGATCATTCACCAAGAGAGTGGCTTAAGTAAATGCAACCTGCTGCCAC 1319  
Db 1261 CTGTGCGAATGATGATCTTTTCAAAAAAAGTAGCGAAAGTCATTGAAACCTGCTGTCAT 1320  
Qy 1320 ATCTAA 1325  
Db 1321 ATTTAA 1326

RESULT 5  
US-10-334-672-4  
; Sequence 4, Application US/10334672  
; Publication No. US20030157646A1  
; GENERAL INFORMATION:  
; APPLICANT: Lanahan, Mike  
; APPLICANT: Koepf, Edward  
; TITLE OF INVENTION: Microbially-Expressed Thermotolerant Phytase For Animal Feed  
; FILE REFERENCE: SYNG-P01-001  
; CURRENT APPLICATION NUMBER: US/10/334,672  
; PRIOR FILING DATE: 2002-12-30  
; PRIOR APPLICATION NUMBER: 60/344,523  
; PRIOR FILING DATE: 2001-12-28  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 1281  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: No. US20030157646A19X gene  
US-10-334-672-4

Query Match 18.3%; Score 242; DB 15; Length 1281;  
Best Local Similarity 52.5%; Pred. No. 3.7e-68;  
Matches 628; Conservative 0; Mismatches 555; Indels 13; Gaps 4;  
Qy 93 ATCGGGCTATACCTTTAGAACGTGTGGTATTTTGGAGTCGCCATGGGTGCTCGCGCGAC 152  
Db 39 ACCAGAAATTGAAGTTGGAATCTGTGTCAATTTGCTCCAGACACGGTGTAGAGCTCCAAC 98  
Qy 153 CAA-ACAAACGAGCTTATGATGATGTTACGCGAGATAAGTGGCCCAATGGCGCGTAAA 211

Db 99 TAAGGCTACTCAGTTGATGCAAGATGTTACTCCAGATGCTTGGCTACTCTACCTGGCTGTAA 158  
Qy 212 AGCGGGTATTTAACGCCACGTTGGTGGGAGTTGGTTCACATTTGATGGGGGGTTTATGG 271  
Db 159 GTTGGGTGAATTTGACTTCAAGAGTGGTGAATTTGCTTACTTGGGTCACTACTCGAG 218  
Qy 272 TGATTACTTTCCGAGCCCTTGGTTTGT - - - AGCGGGGGGATGCTCCGGCAGAGGGGGGGT 328  
Db 219 ACAAAGATTGGTTGCTGATGTTTGGTCCAAAGTGGTGTGCTCCACAATCTGGTCAAGT 278  
Qy 329 ATATGCAAGCGAGATTCGATCAACGTTACCGGCTTAACCGGACAGGCATTTCTTGATGG 388  
Db 279 TGTCTATCATTTGCTGATTTGATGAAAGAACTAGAAAGACTGGTGAAGCCTTCGCTGCCG 338  
Qy 389 TGTGGCTCCGGGGTGGTTTGACCGTGCAATAATCAGGCCGATTTGAAAAAGACCCGATCC 448  
Db 339 TTTGGCCCCAGACTGTGCTATCACTGTTCACTCACTCACTGATATTTCTCTCCAGATCC 398  
Qy 449 CTTGTTCCATCCGGTAGAGACTGCGGTGTGTAGTTAGAACACGCAACAGATTAAGC 508  
Db 399 ATTGTTCAACCCCATTTGAAGACTGCTGTCTGCTCAATTGGATAACGCTAACCTTCTGATGC 458  
Qy 509 GATTGAAGAACGTTGGCGGGCGGTTAGATACGTTGAGCCAGCGCTACGCTAAACCTTT 568  
Db 459 CATCTTGGAAAGAGCTGGTGTCTATCGTGTCTCACTGCTCACTACCAAGCTGCTT 518  
Qy 569 TGCCAGATGGGGACGTTGTAATTTTGGCGCTTCTCTTATTTGCAAAATCTTTTGAACA 628  
Db 519 CAGAGAAATGCAAGAGCTCTTGAATTTCCCAACAATTAATCTGTGTTTGAAGAGAGAGAA 578  
Qy 629 GCAGGAAACGCTGTGATTTTGCCCACTTTGGCGCAATGAAAGTAAACGTTAATAAGA 688  
Db 579 GCAAGAGCAATCTTTGCTTGTGCTCAAGCTTGCATCTGATTTGAGGTCTCTGCTGA 638  
Qy 689 AGGACAAAAGTGAACCTCAGTGGGCCACTGGGCTTATCATGACATTTGGGTGAATCTT 748  
Db 639 TTG - - - - TGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 692  
Qy 749 CTTATTACAAACGCAAGCCATGCGAGAGTTGCTGCGCAACGGCTAAAGGGCGGA 808  
Db 693 CTTGTTGCAACAGCTCAAGGTATGCCAGAACCGAGTTGGGTGAGATCACTGATTTCTCA 752  
Qy 809 GAATTGGGTATCTTTATTGCTTACATAACGGGCAATTTAATTAATGSCAAACACCC 868  
Db 753 CCATGGAACACCTTTGTTGCTTGCACAAAGCTCAATTCGATTTGCTGACAGAACTCC 812  
Qy 869 GTATATCGCCGCTATTAAGGAGCGCATTTATACAGCAGATAGATACGGTTTAAACCT 928  
Db 813 AGAAGTCCGTAGATCCAGAGCTACTCCATTTGTTGACTTGTATCAAGACCGCTTTGAC - - - 869  
Qy 929 TCAACTGGATGCTCAGGGGCAAAAGCTACCCATTTAGCCCAAAACCGGGTCTTTGCTCT 988  
Db 870 TCCACCCCAACAGAGCAAGCTTACGGTGTACCTTGCCAACTTCTGCTTGTTCAT 929  
Qy 989 CGGTGGGATGATACCAATATTTGCCAATATTTGGGGTATGCTGGGAGCCGAGCTGGCAGCT 1048  
Db 930 TGCGGTCAAGATTAACCTTGGCTAACTTTGGGTGGTGGCTTGGAAATGGAACCTGACCTT 989  
Qy 1049 ACCGAGCAACCTGATTAATCTCCACAGGTGGGGGATTTGGTTTGAACATATGCGCAAAA 1108  
Db 990 GCGAGTCAACAGATTAACCTCCACAGGTGGTGAATTTGGTCTTCGAAAGATGCGCTCG 1049  
Qy 1109 TCCGATTAACCAACAGCGCTACGTTGGGTTGAAGATGTTCTTCAACAGATGATGATGAT 1168  
Db 1050 ACTGCTGTGATAACTCTCAATGGAATTCAGTCTCCTGCTCTTCCAAACCTTTGCAACAAAT 1109  
Qy 1169 GCGTAATCCCGAAATTTGGATCTGAAAAATAAACCCAGCGGGTATTTTCCGTTGCGAGT 1228  
Db 1110 GAGAGCAAGACTCCATTTGCTTGAACACTCCACAGGTGAAGTCAAGTTGACCTTGGC 1169  
Qy 1229 TGCTGGTTGTGAAATAACCGTGAAGCTTAAGCTTTGGAGCTTTGAGATTTGCCAAA 1284  
Db 1170 TGGTTGTGAAGAAAGAACGCTCAAGGTATGTTGTTCTTTGGCTGGTTTCACTCAA 1225

RESULT 6

US-10-334-671-4  
 ; Sequence 4, Application US/10334671  
 ; Publication No. US20030170293A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lanahan, Mike  
 ; APPLICANT: Koepf, Edward  
 ; APPLICANT: Kretz, Keith  
 ; TITLE OF INVENTION: Microbially-Expressed Thermotolerant Phytase For Animal Feed  
 ; FILE REFERENCE: 70098  
 ; CURRENT APPLICATION NUMBER: US/10/334,671  
 ; CURRENT FILING DATE: 2002-12-30  
 ; PRIOR APPLICATION NUMBER: 60/344,523  
 ; PRIOR FILING DATE: 2001-12-28  
 ; NUMBER OF SEQ ID NOS: 4  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 4  
 ; LENGTH: 1281  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: No. US20030170293A19X gene  
 US-10-334-671-4

Query Match 18.3%; Score 242; DB 15; Length 1281;  
 Best Local Similarity 52.5%; Pred. No. 3.7e-68;  
 Matches 628; Conservative 0; Mismatches 555; Indels 13; Gaps 4;

QY	93	ATCGGGCTATCTTTAGAACGTTGGTTATTTTGGTGGCCATGTTGGTGGCCGAC	152
DB	39	ACCAGAAATGAAGTTGAATCTGTTGCTTCTCCAGACACGGTGTAGAGCTCCAC	98
QY	153	CAA-ACACGAGCTTATGAATGTTACGCCAGATAGTGGCCACAAATGGCCGTA	211
DB	99	TAAGGCTACTCAGTTGATGCAAGATGTTACTCCAGATGCTTTGGCCTACCTGGCCTGTAA	158
QY	212	AGCGGGTATTAAACGACCTGGTGGGAGTGGTCACTGATGGGGGTTTATGG	271
DB	159	GTGGGTGAATGATCTCCAGAGGTTGTAATGATGCTTACTTGGGTCACTACTGAG	218
QY	272	TGATTAATCTTCGACGCTTGGTTGTTT---AGCGCGGGATGTCGGCAGAGGGGGGT	328
DB	219	ACAAAGATGTTGCTGATGTTTGTGTCCTCCAAAGTGGTGTCCACATCTGGTCAAGT	278
QY	329	ATATGCACAGGAGATATCGATACGACGTTACCGGCTTAACCGCAGAGGATTTCTGTATG	388
DB	279	TGCTATCATTTGCTGATGTTGATGAAGAACTAGAAAGACTGGTGAAGCCTTTCGTCGGG	338
QY	389	TGTGCTCCGGGTGTGGTTGACCGTGCATAATCAGCGCGAATTTGAAAAGACCGATCC	448
DB	339	TTTGGCCCCAGACTGTGCTATCCTGTTTACACTCAAGCTGATCTTCTCTCCAGATCC	398
QY	449	CTGTTTCCATCCGGTAGAGACTGGCGTGTGAAGTTAGACACGCAACAGATAAAGC	508
DB	399	ATTGTTCAACCCANTTGAAGACTGGTGTCTGTCATTTGGATTAACGCTAACGTTACTGATC	458
QY	509	GATTTGAAGAACGATTTGGCGGCGCTTGAATACGGTGGAGCGGCTACGCTAAACCTTT	568
DB	459	CATCTTGAAGAGCTGGTGGTTCTATCGCTGACTTCTGCTGCTACCTACCAACTGCTTT	518
QY	569	TGCCAGATAGGGGAGCTGTGAATTTTGGCGCTTCTCTCTTATTTGCAAAATCTTTGCAACA	628
DB	519	CAGAGATTTGGAAGAGCTTTGAATCTCCCAATCTAACCTTGTGTTTGAAGAGAGAGAA	578
QY	629	GCAAGGAAACGCTGTGATTTTGGCCACTTTTGGCGCCCAATGAAGTTAACGTTTAAATAAGA	688
DB	579	GCAAGACGAATCTTGTCTTCTGACTCAAGCCTTGCATCTGAATTTGAAGTCTCTGTGA	638
QY	689	AGGGAACAAAGTACCTCAGTGGGCCACTGGCGGTATCATCGCATTTGGGTGAATCTTT	748
DB	639	TTG-----TGTCCTCTGACTGGTGTCTCTCTCTTGGCTTCTATGTTGCTGAATCTTT	692

RESULT 7

US-10-156-660-1  
 ; Sequence 1, Application US/10156660  
 ; Publication No. US20030103958A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Short, Jay M.  
 ; APPLICANT: Kretz, Keith  
 ; APPLICANT: Gray, Kevin A.  
 ; APPLICANT: Barton, Nelson R.  
 ; APPLICANT: Garrett, James B.  
 ; APPLICANT: O'Donoghue, Eileen  
 ; APPLICANT: Mathur, Eric J.  
 ; TITLE OF INVENTION: PHYTASES, NUCLEIC ACIDS ENCODING THEM  
 ; TITLE OF INVENTION: AND METHODS FOR MAKING AND USING THEM  
 ; FILE REFERENCE: 09010-029007  
 ; CURRENT APPLICATION NUMBER: US/10/156,660  
 ; CURRENT FILING DATE: 2002-10-01  
 ; PRIOR APPLICATION NUMBER: US 09/866,379  
 ; PRIOR FILING DATE: 2001-05-24  
 ; PRIOR APPLICATION NUMBER: US 09/580,515  
 ; PRIOR FILING DATE: 2000-05-25  
 ; PRIOR APPLICATION NUMBER: US 09/318,528  
 ; PRIOR FILING DATE: 1999-05-25  
 ; PRIOR APPLICATION NUMBER: US 09/291,931  
 ; PRIOR FILING DATE: 1999-04-13  
 ; PRIOR APPLICATION NUMBER: US 09/259,214  
 ; PRIOR FILING DATE: 1999-03-01  
 ; PRIOR APPLICATION NUMBER: US 08/910,798  
 ; PRIOR FILING DATE: 1997-08-13  
 ; NUMBER OF SEQ ID NOS: 4  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 1  
 ; LENGTH: 1308  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence

QY	749	CTTATTACAAACGACACAGCCATCCAGAGGTTGCTGGCAACCGCTAAAGGGCGGA	808
DB	693	CTTGTTCGCAACAGCTCAAGGTATGCCAGAACAGGTTGGGTAGAACTCATGATCTCA	752
QY	809	GAATTGGGTATCTTATTGTCTATACATAACGCGCAATTTAATTTAATGGCAAAACACC	868
DB	753	CCAAATGGAACACCTTGTGCTCTTGCACAAACGCTCAATTCGATTTGCTGCAGAGAACTCC	812
QY	869	GTATATCGCCCTCTATAAGGACGCCATTATTACAGCAGATAGATACGGCTTTAACCTT	928
DB	813	AGAAATCGGTAGATCCAGAGCTACTCCATTTGTGGACTTTGATCAAGACCGCTTTGAC---	869
QY	929	TCAACTGATGCTCAGGGGCAAAAGCTACCCATTTTCAGCCCAAAACCGGGTCTTGTCTT	988
DB	870	TCCACACCCACACAGAACGCTTACGGTGTACTTACCTTGCACACTTCTGTCTTGTTCAT	929
QY	989	CGGTGGCATGATACCAATATTGCCAATATTGCCGGTATGCTGGGAGCCGACTGGCAGCT	1048
DB	930	TGCCGGTACGATACCTTAACCTTGGCTTAACCTTGGGTGGCTTGGAAATGGAATGGACCTT	989
QY	1049	ACCCGAGCAACCTGATAATPACTCCACAGGTGGGGGATTTGGTTTGAACATATGGCAAAA	1108
DB	990	GCCAGGTCAACAGATAACACCTCCACAGGTGCTGAATTTGGTCTTCGAAAGATGGCGTCG	1049
QY	1109	TCCGATTAACCAACGAGCTGCTGTTGCGGTGAAGATGTTTACCAAAACGATGATGATGTT	1168
DB	1050	ACTGCTGATACTCTCAATGATTAAGTCTCCTTGGTCTTCCAAACCTTGCACAAAT	1109
QY	1169	GGCTAATGCCGAAAATTTGATCTGAAAATAAACCCAGCGGGTATTTATTTCCGTTGCAAT	1228
DB	1110	GAGAGACAGACTCCCATTTGCTTGAACACTCCACAGGTGAAGTCAAGTTGACCTTGGC	1169
QY	1229	TGCTGGTGTGAAAATAACGGTGAAGATGTTTGGAGCTTTGGAGCTTGATACATTTCCAAA	1284
DB	1170	TGTTGTGGAAGAAGAAACGCTCAAGGTATGTTCTTTGGCTGGTTTCTTCACTCAAA	1225

FEATURE:  
; OTHER INFORMATION: modified phytase enzyme  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)... (1308)  
US-10-156-660-1

Query Match 17.1%; Score 226.2; DB 15; Length 1308;  
Best Local Similarity 53.3%; Pred. No. 5.9e-62;  
Matches 571; Conservative 0; Mismatches 489; Indels 13; Gaps 4;

Qy	107	TAGAACGTGGTGTATTTGAGTCGCCCATGTTGCTCGCCGACCAA-ACAACGCAGC	165
Db	89	TGGAAGTGGTGTATTTGAGTCGCCCATGTTGCTCGCCGACCAA-ACAACGCAGC	148
Qy	166	TTATGATGATGTACCCGAGTAAGTGGCCACATGCGCGGTAAAGCGGGGTATTAA	225
Db	149	TGATGAGGATGTACCCCGAGATGCGCCACATGCGCGGTAAAGCGGGGTATTAA	208
Qy	226	CGCCACGTGGTGGGAGTGGTGCATGATGGGGGGTATTTATGGTGATTTACTTCGCA	285
Db	209	CACCGCGGTGGTGAGCTAATCGCTATCTCGACATTAATGGCGTCAGCGTCTGGTAG	268
Qy	286	GCCTTGGTGTAGCG---GCGGGATGTCCGGCAGAGGGGGGTATATGCAAGGCAG	342
Db	269	CCGACGGATTTGCTGCTTAATGTTGGTGGCGCCGAGTCTGGTCAGGTGCGGATTTGCTG	328
Qy	343	ATATCGATCAACGTACCGGCTTAACCGGACAGGCATTTCTTGATGGTGTGGCTCCGGGT	402
Db	329	ATGTCGACGAGGTACCGGTAACAGCGGAGCGCTTCGCGCGGGGTGGACCTGACT	388
Qy	403	GTGTTTGACGTGCATATACGCGCATTTGAAAGACGATCCCGTTCCTCCATCCGG	462
Db	389	GTGCAATTAACCGTACATACCGGAGATACGTCCTCGATCCCGATTCGTTATTAATCCTC	448
Qy	463	TAGAGACTGCGGTGTGAAGTTAGACAAACGACAAACAGATAAGCGATTAAGACGAT	522
Db	449	TAAAACTGGGTTTGCACTGATTAACCGGACGAGTACGCGATCTCTCGAGGG	508
Qy	523	TGGCGCGCGTTAGATACGGTGAAGCGAGCGCTAAGCTAAACCTTTTGCCCGAGTGGGG	582
Db	509	CAGGAGGCTCAATGCTGACTTTACCGGGCATTTCAACCGCGGTTTGGCGAACTGGAAC	568
Qy	583	ACGTGCTGAATTTGGCGTCTCTTATTGCAAACTTTTGCAACGACGAGGAAACGT	642
Db	569	GGGTGCTTAATTTCCGCAATCAAACTTGTCCTTAAACGTGAGAAACGAGCGAAGCT	628
Qy	643	GTGATTTGCCCACTTTGCGGCCAATGAAGTTAAAGTTAATAAGAGGGACAAAGTGA	702
Db	629	GTTCATTAAACGACGAGCATTAACATCGAACTCAAGGTGACGCGGACTGTCTCATTA	688
Qy	703	CCCTCAGTGGCCACTGGCGTTATCATCGACATTTGGTGAATCTTCTTATTACAAACG	762
Db	689	CC-----GTTGCGGTAAAGCTTCGATCAATGCTGACGGAGATATTTCTCTGCAACA	742
Qy	763	CACAAGCCATGCCAGAGTTGCTTGGCAACGGCTTAAAGGGGGGAGAAATTTGGGTATCCT	822
Db	743	CACAGGGAATGCCGAGCCGGGTGGGAAGGATCACCGATTACACAGTGAACACCT	802
Qy	823	TATTGTCATTACATAACGCGCAATTTAATTAAATGGAACAAACACCGTATATCGCCGTC	882
Db	803	TGCTAAGTTTGCATACGCGCAATTTGATTTGCTACAAACGACGCGGAGGTTGCGCGCA	862
Qy	883	ATAAAGGAGCGCCATTATTACAGCATAGATACGCGCTTTAACCTTCACTGATGTC	942
Db	863	CGCGCGCCACCCGTTATTAGATTGATCAAGACAGGTTGACGCCCATC---CACCGC	919
Qy	943	AGGGGCAAAAGCTACCCATTTCAGCCCAAAACCGGGTCTTTGTTCTCGTGGGCATGATA	1002
Db	920	AAAAACAGGCGTATGGTGTGACATTAACCACTTCAGTGTCTTTTATCGCGGACAGATA	979
Qy	1003	CCAAATTTGCCAATTTGCGGGTATGCTGGAGCGAGCTACCGGACACCTTG	1062

Db 980 CTAATCTGGCAAAATCTCGCGCGGCACCTGGAGCTCAACTGGAGCTTCCCCTGACGCGG 1039  
Qy 1063 ATAATACTCCACAGGTGGGGATTTGGTTTTTGAATATGCAAAATCCGGATTAACCAACC 1122  
Db 1040 ATAACACGCGCCAGGTGGTGAATGTTGTTGAACGCTGGCGCTAAGCGGATAACA 1099  
Qy 1123 AGCGGTACTTGGCGGTGAAGATGTTCTACCAAAACGATGATGATGCTGAT 1175  
Db 1100 GCCAGTGATTCAGGTTTCGCTGGTCTTCCAGACTTTTACAGCAGATGCGTAT 1152

RESULT 8  
US-09-866-379-5  
; Sequence 5, Application US/09866379  
; Patent No. US20020136754A1  
; GENERAL INFORMATION:  
; APPLICANT: DIVERSA CORPORATION  
; APPLICANT: SHORT, Jay  
; APPLICANT: KRETZ, Keith  
; APPLICANT: GRAY, Kevin  
; APPLICANT: BARTON, Nelson  
; APPLICANT: GARRETT, James  
; APPLICANT: O'DONOGHUE, Eileen  
; TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYTASES AND USES THEREOF  
; FILE REFERENCE: DIVER1370-7  
; CURRENT APPLICATION NUMBER: US/09/866.379  
; CURRENT FILING DATE: 2001-05-24  
; PRIOR APPLICATION NUMBER: US 09/580,515  
; PRIOR FILING DATE: 2000-05-25  
; PRIOR APPLICATION NUMBER: US 09/318,528  
; PRIOR FILING DATE: 1999-05-25  
; PRIOR APPLICATION NUMBER: US 09/291,931  
; PRIOR FILING DATE: 1999-04-13  
; PRIOR APPLICATION NUMBER: US 09/259,214  
; PRIOR FILING DATE: 1999-03-01  
; PRIOR APPLICATION NUMBER: US 08/910,798  
; PRIOR FILING DATE: 1997-08-13  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 5  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(1901)  
; OTHER INFORMATION: n is any nucleotide  
US-09-866-379-5

Query Match 17.0%; Score 225.2; DB 9; Length 1901;  
Best Local Similarity 53.2%; Pred. No. 1.6e-62;  
Matches 571; Conservative 0; Mismatches 489; Indels 13; Gaps 4;  
Qy 107 TAGAACGTGGTGTATTTGAGTCGCCCATGTTGCTCGCCGACCAA-ACAACGCAGC 165  
Db 276 TGGAAGTGGTGTATTTGAGTCGCCCATGTTGCTCGCCGACCAA-ACAACGCAGC 335  
Qy 166 TTATGAATGATGTACGCCAGATTAAGTGGCCACAAATGCGCGGTAAAGCGGGGTATTAA 225  
Db 336 TGATGAGGATGTACCCCGAGATGCGCCAGACGATGCGCGGTAAAGCGGGGTATTAA 395  
Qy 226 CGCCACGTGGTGGGAGTGGTGCATGATGGGGGGTATTTATGGTGTATTCTGCA 285  
Db 396 CACCGCGGTGGTGGAGCTAATCGCTATCTCGACATTAACCAACGCGCGGTCTGGTAG 455  
Qy 286 GCCTTGGTGTAGCG---GCGGGATGTCCGGCAGAGGGGGGTATATGCAAGGCAG 342  
Db 456 CCGACGATTTGCTGGCGAAAAAGGGCTGCCCGCAGTCTGGTCAGTCCGATTATTGCTG 515  
Qy 343 ATATCGATCAACGTCACCGCTTAACCGGACAGGATTTCTTGTGATGGTGTGCTCGGGGT 402  
Db 516 ATGTCACAGCGATACCCGTAACAGCGGAGCGCTTCGCCCGCGGGCTGGCACTGACT 575

QY 403 GTGGTTTACCGTGCATATCAGGCGGATTTGAAAAGACCGATCCCTGTTCATCCGG 462  
DB 576 GTGCAATAACCGTACATACCCAGGAGAGATCGTCCAGTCCCGATCCGTATTATATCCTC 635  
QY 463 TAGAGACTGGCGTGTGAAGTAGACACGACACAGATAAGAGGATTTGAAGACGAT 522  
DB 636 TAAAAACTGGCGTTTGGCAACTGAGATACCGGACGAGTACTGACGCGCATCTCAGACGG 695  
QY 523 TGGCGCGCGCTTAGATACGCTGAGCAGCGCTACGCTAAACCTTTTCCCGAGATGGGG 582  
DB 696 CAGGAGGCTCAATTGCTGACTTTACCGGACATCGGCAACCGCGTTTCGCGAATCGAAC 755  
QY 583 AGTGTCTGAATTTTGGCGCTCTCTCTTATTCGCAATCTTTGCAACAGAGGAAAAAGCT 642  
DB 756 GGGTGTCTTAATTTTCCGCAATCAAACTTTGTCTTAAACGCTGAGAAACAGGACGAAAGCT 815  
QY 643 GTGATTTTGGCCACTTTTGGCGCAATGAAGTTAACTTAAATAAGAGGACAAAGTGA 702  
DB 816 GTTCATTAACGAGCATATCACTCGAATCTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 869  
QY 703 CCCTCAGTGGCGACTGGCGTCTCTCTTATTCGCAATCTTTGCAACAGAGGAAAAAGCT 762  
DB 870 CATTAACCGGCTGAGTAACTCGCATCAATGCTGACGAGATATTTCTCTGCAACAG 929  
QY 763 CACAAGCCATCCAGAGGTTGCTGCGCAACCGCTAAAGGGGCGGAGATTTGGGTATCCT 822  
DB 930 CACAGGGAATGCGGAGCGCGGTTGGGAGAGATCAAGATTACACAGTGGAAACCT 989  
QY 823 TATTGTCTATTAATAACCGCAATTTAATTTAATGGCAAAACACCGTATATCGCGCTC 882  
DB 990 TGTAAAGTTTGCATAACCGCAATTTTATTTGCTACAGCGACCGCAGAGGTTGCCCGCA 1049  
QY 883 ATAAAGGAGCGCATTTATCAGCAGATAGATCGGCTTTAAACCTTCACTGATGCTC 942  
DB 1050 GCGCGGCCACCCGTTTATGATTTGATCAAGACAGCGTTGACGCCCATC---CACCGC 1106  
QY 943 AGGGGCAAAAGCTACCCATTTACGCCAAAACCGGCTTTGTTCTCGGTGGGCAATGATA 1002  
DB 1107 AAAAAGCAGCGTATGCTGATGATTAACCATCTCAGTCTGTTTATCGCGGACAGGATA 1166  
QY 1003 CCAATATTGCAATATTGCGGATGCTGGAGCGGACTGCGAGTACCGGAGCACTG 1062  
DB 1167 CTAATCTGGCAATCTGCGCGGCGACTGAGCTCACTGACCGCTTCCCGGTGAGCGCG 1226  
QY 1063 ATAATCTCCACCGAGTGGGGAATGTTTGAACCTATGCAAAATCCGAGATAACCAAC 1122  
DB 1227 ATACACGCGCGAGTGTGAATGTTGATGTTGATGCTGCGCTGAGCGATACCA 1286  
QY 1123 AGCGTACGTTGCGGTGAAGATGTTTACCAAAAGATGATGATGCTGCTTAAT 1175  
DB 1287 GCCAGTGAATCAGGTTTTCGCTGCTTCCAGACTTTTACAGCAGATGCTGAT 1339

## RESULT 9

US-09-866-379-6  
; Sequence 6, Application US/09866379  
; Patent No. US20020136754A1  
; GENERAL INFORMATION:  
; APPLICANT: DIVERSA CORPORATION  
; APPLICANT: SHORT, Jay  
; APPLICANT: KRETZ, Keith  
; APPLICANT: GRAY, Kevin  
; APPLICANT: BARTON, Nelson  
; APPLICANT: GARRETT, James  
; APPLICANT: O'DONOGHUE, Eileen  
; TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYTASES AND USES THEREOF  
; FILE REFERENCE: DIVER1370-7  
; CURRENT APPLICATION NUMBER: US/09/866,379  
; PRIORITY FILING DATE: 2001-05-24  
; PRIORITY APPLICATION NUMBER: US 09/580,515  
; PRIORITY FILING DATE: 2000-05-25  
; PRIORITY APPLICATION NUMBER: US 09/318,528  
; PRIORITY FILING DATE: 1999-05-25

; PRIOR APPLICATION NUMBER: US 09/291,931  
; PRIOR FILING DATE: 1999-04-13  
; PRIOR APPLICATION NUMBER: US 09/259,214  
; PRIOR FILING DATE: 1999-03-01  
; PRIOR APPLICATION NUMBER: US 08/910,798  
; PRIOR FILING DATE: 1997-08-13  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 6  
; LENGTH: 1901  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(1901)  
; OTHER INFORMATION: n is any nucleotide  
US-09-866-379-6

Query Match 17.0%; Score 225.2; DB 9; Length 1901;

Best Local Similarity 53.2%; Pred. No. 1.6e-62;  
Matches 571; Conservative 0; Mismatches 489; Indels 13; Gaps 4;

QY 107 TAGAAGCTGTGCTTATTTTGGTCCCATGTTGCTCGCGACCAA-ACAAAGCGAGC 165  
DB 276 TCGAAGTGTGCTGATTGCTGCTCATGCTGCTGCTTCAACCAAGCGCAGCAAC 335  
QY 166 TTATGAATGATTTAGCCAGATAAGTGGCCCAATGCGCGGTAAAGCGGGGTATTTAA 225  
DB 336 TGATGAGGATGTCAACCCAGACGATGCGCAACCTGCGCGGTAAAGCTGGGTTCGCTGA 395  
QY 226 CCGCAGCTGCTCGGAGTTGGTCAATGATGGGGGGTTTTATGCTGATTTACTTTGCGCA 285  
DB 396 CACCGGGGTTGCTGAGCTAATCGCTATCTCGGACATACCTGCGCGCAGCTCTGGTAG 455  
QY 286 GCGTGGTGTGTTAGCG--GCGGATGTCCGCGAGAGGGGGGTATATGACAGGCGAG 342  
DB 456 CCGACGATTTGCTGCGAAAAAGGCTGCGCGAGTCTGCTGCTGCTGCTGCTGCTGCTG 515  
QY 343 ATATCGATCAAGTACCGCTTAAACCGGACAGGCAATTTCTGATGCTGCTGCTGCTGCTG 402  
DB 516 ATGTGACAGCGCTTACCGGTAAACAGGCGAGGCTTTCGCCCGCGGCTGGACCTGACT 575  
QY 403 GTGGTTTGAACCGTCAATATCAGGCGGATTTGAAAAGACCGATCCCTGTTTCCATCCGG 462  
DB 576 GTGCAATAACCGTACATACCCAGGAGATACGTCAGTCCGATCCGTTATTTAATCCTC 635  
QY 463 TAGAGACTGGCGTGTGAAGTTAGACAGGACACAGATTAAGCGATTGAAGAACGAT 522  
DB 636 TAAAAACTGGCGTTTGGCAACTGGATACGCGAAGCTGACGCGATCTTCAGCAGGG 695  
QY 523 TGGCGGGGCGCTTAGATACGCTGAGCAGCGCTACGCTAAACCTTTTCCCGAGATGGGG 582  
DB 696 CAGGAGGCTCAATTGCTGACTTTACCGGCAATCGGCAACCGCGTTTCGCGAATCGAAC 755  
QY 583 AGTGTCTGAATTTTGGCGCTTCTCTTATTCGCAATCTTTGCAACAGGAGGAAAAAGCT 642  
DB 756 GGGTGTCTTAATTTTCCGCAATCAAACTTTGCTGCTTAAACGCTGAGAAACAGGAAAGCT 815  
QY 643 GTGATTTTGGCCACTTTGGGCGCAATGAAGTTAACTTAAATAAGAGGACAAAGTGA 702  
DB 816 GTTCATTAACGAGCATTTACCATCGAATCTCAAGTGGAGCGCGCA-----CAATGCT 869  
QY 703 CCCTCAGTGGCGACTGGCGTATCATGACATTTGGGTGAATCTTTTATTTACAAAAG 762  
DB 870 CATTAACCGGCTGAGTAAAGCTCGCATCAATGCTGACGAGATATTTCTCTCGCAACAG 929  
QY 763 CACAAGCCATGCGAGGTTGCTGCGCAACCGCTAAAGGGGCGGAGATTTGGGTATCCT 822  
DB 930 CACAGGGAATGCGGAGCGCGGTTGGGAGAGATCAACGATTCAACAGTGGAAACCT 989  
QY 823 TATTGTCTATTAATAACCGCAATTTAATTTAATGGCAAAACACCGTATATCGCGCTC 882  
DB 990 TGTAAAGTTTGCATAACCGCAATTTTATTTGCTACAGCGACCGCAGAGGTTGCCCGCA 1049



QY 883 ATAAAGGAGCGCCATTATTACAGCAGATAGATACGGCTTTTAAACCTTTCAACTGGATGCTC 942  
Db 1050 GCGCGCCACCCCGTTATTAGATTGATCAAGACAGCGTTTGAAGCCCATC---CACCGC 1106  
QY 943 AGGCGCAAAAGCTACCCATTTCAGCCCAAAACCGGGTCTTTCTCGGTGGGCATGATA 1002  
Db 1107 AAAACAGGCGTATGGTGTGACATTACCTTCACTTCACTGCTGTTTATCGCGGACAGATA 1166  
QY 1003 CCAATATTGCCAATATTTCGGGTATGCTGGAGCCGAGCTGGCAGTACCGGAGCAACCTG 1062  
Db 1167 CTAATCTGGCAAAATCTCGCGGGGCGCACTGGAGCTCAACTGAGCGCTTCCCGGTGAGCCG 1226  
QY 1063 ATAATCTCCACAGGTGGGGATTGGTTTTCGAATATGCGCAAAATCCGGATAACCC 1122  
Db 1227 ATACAGCGCGCAGGTGGTGAAGTGTGTTGAAAGCTGGCTGCGCTAAGCGATAACA 1286  
QY 1123 AGCGCTACGTGGCGTGAAGATGTTTACCAACGATGGATCAGTTCGCTAAT 1175  
Db 1287 GCCAGTGGATTGAGTTTCGCTGCTCTCCAGACTTTACAGCAGATGCGTGAT 1339  
RESULT 10  
US-09-866-379-7  
; Sequence 7, Application US/09866379  
; Patent No. US20020136754A1  
; GENERAL INFORMATION:  
; APPLICANT: DIVERGA CORPORATION  
; APPLICANT: SHORT, Jay  
; APPLICANT: KRETZ, Keith  
; APPLICANT: GRAY, Kevin  
; APPLICANT: BARTON, Nelson  
; APPLICANT: GARRETT, James  
; APPLICANT: O'DONOGHUE, Eileen  
; TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYTASES AND USES THEREOF  
; FILE REFERENCE: DIVER1370-7  
; CURRENT APPLICATION NUMBER: US/09/866,379  
; CURRENT FILING DATE: 2001-05-24  
; PRIOR APPLICATION NUMBER: US 09/580,515  
; PRIOR FILING DATE: 2000-05-25  
; PRIOR APPLICATION NUMBER: US 09/318,528  
; PRIOR FILING DATE: 1999-05-25  
; PRIOR APPLICATION NUMBER: US 09/291,931  
; PRIOR FILING DATE: 1999-04-13  
; PRIOR APPLICATION NUMBER: US 09/259,214  
; PRIOR FILING DATE: 1999-03-01  
; PRIOR APPLICATION NUMBER: US 08/910,798  
; PRIOR FILING DATE: 1997-08-13  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 7  
; LENGTH: 1901  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)-(1901)  
; OTHER INFORMATION: n is any nucleotide  
US-09-866-379-7  
Query Match 17.0%; Score 225.2; DB 9; Length 1901;  
Best Local Similarity 53.2%; Pred. No. 1.6e-62;  
Matches 571; Conservative 0; Mismatches 489; Indels 13; Gaps 4;  
QY 107 TAGAAGCTGGTATTATTGAGTCGCGCATGGTTCGCTCGCGGACCAA-ACAAAGCAGC 165  
Db 276 TGAAGAGTGGGTGATGTCAGTCGTCATGGTGTGCTCTCAACCAAGGCGCAGCAAC 335  
QY 166 TTATGAATGATGTTAGCCAGATAAGTGGCCAAATGGCGCGTAAAGCGGGTATTTAA 225  
Db 336 TGATGAGGATGTCACCCAGAGCGCATGGCCACCTTGGCGGGTAAACTGGGTGGCTGA 395  
QY 226 CGCCAGCTGGTGGAGTGGTGCACATTAAGTGGGGGGGGTTTATTAGTGATTAATTCCGA 285

Db 396 CACCGCGNGTGGTGAAGCTAATCGCTATCTCGGACATTTACCAAGCCGAGCGTCTGGTAG 455  
QY 286 GCCTTGGTTTGTAGCG--GCGGATGTCGCGAGAGGGGGGGGTATATCCACAGGCGAG 342  
Db 456 CCGACGAGTTCCTGGCGAAAAGGGCTGCCCGCAGTCTGGTCAGGTCCGATTTATTGTG 515  
QY 343 ATATCGATCAACGTATCCGCTTAAACCGGACAGGCAATTTCTTGTATGTTGGTCTCGGGGT 402  
Db 516 ATGTCGAGAGCGTACCGGTAAAAACAGGCGAAGCCTTTGCGCGCGGGGTGGCAGCTGACT 575  
QY 403 GTGGTTTACCGCTGATATATCAGCCGATTTGAAAAAGACCGATCCCTCTTTCATCCGG 462  
Db 576 GTGCAATTAACGTATACCCAGGAGATAGCTCAGTCCCGATCCGTTATTTATTCCTC 635  
QY 463 TAGAGCTGGCGGTGTAAAGTTAGCAACGACCAAGATATAAGACGATTAAGAAACGAT 522  
Db 636 TAAAACTGGCGTTTCCCACTGGATAACGCGAACGTAAGTCTGAGCGGATCTCAGCAGGG 695  
QY 523 TGGCGGGCGCTTAGATACGGTGGAGCGGCTAGCTAAACCTTTTCCCGAGATGGGG 582  
Db 696 CAGGAGGTCAATTGCTGACTTTACCGGCGCATCCGCAAAACCGCGTTTTCGCGAATCGAAC 755  
QY 583 ACCTGCTGAATTTTGGCGCTTCTCTTATTGCAAAATCTTTGCAACAGCAAGAAAGAGT 642  
Db 756 GGGTCTTAAATTTTCGCGCAATCAAACTTGTGCTTAAACGCTGAGAAACAGGACGAAAGCT 815  
QY 643 GTGATTTTGGCCACTTTTGGCGGCAATGAAGTTACCTTATAAAGAGGAGCAAAAGTGA 702  
Db 816 GTTCATTAACGAGCGCATTTACATCGGAATCAAGGTGAGCGCGA-----CAATGTCT 869  
QY 703 CCCTCAGTGGCGCACTGGCGTTATCATCGACATTTGGGTGAAATCTTCTTATTACAAAACG 762  
Db 870 CATTAACCGGTGCGTGAAGCTCGCATCAATGCTGACGAGATATTTCTCTGCAACAG 929  
QY 763 CACAAGCCTGCGCAGAGTTTCCCTGGCAACGCTTAAAGGGGCGGAGAAATGGGTATCCT 822  
Db 930 CACAGGGAATGCCGAGCGCGGGTGGGGAAGGATCACCGATTACACAGGTGGAAACCT 989  
QY 823 TATTGTCAATTACATAAAGCGCAATTTAAATTAATGGCAAAACACCGTATATCCCGCTC 882  
Db 990 TGCTAAGTTTGCATAACGCGCAATTTTATTTGTACAAACGCGCAGAGGTGGCCGCA 1049  
QY 883 ATAAAGGAGCGCCATTATTACAGCAGATAGATACGGCTTTTAAACCTTTCAACTGATGCTC 942  
Db 1050 CCGCGCGCACCCCGTTATTAGATTGATCAAGACAGCGTTTGAACGCGCCCATC---CACCGC 1106  
QY 943 AGGCGCAAAAGCTACCCATTTCAGCCCAAAACCGGGTCTTTCTCGGTGGGCATGATA 1002  
Db 1107 AAAACAGGCGTATGGTGTGACATTACCCACTTCAGTGTCTTTATCGCGGACAGATA 1166  
QY 1003 CCAATATTGCCAATATTTCGGGTATGCTGGAGCGGACTGGCAGCTACCGAGCAACCTG 1062  
Db 1167 CTAATCTGGCAAAATCTCGCGGGCGCACTGGAGCTCAACTGAGCGCTTCCCGGTGAGCCG 1226  
QY 1063 ATATATCTCCACAGGTGGGGATTGGTTTTCGAATATGCGCAAAATCCGGATAACCC 1122  
Db 1227 ATAAACGCGCGCAGGTGGTGAAGTGTGTTGAACGCTGGCGTGGCGTAAAGCGATAACA 1286  
QY 1123 AGCGCTACGTGGCGTGAAGATGTTTACCAACGATGGATCAGTTCGCTAAT 1175  
Db 1287 GCCAGTGGATTGAGTTTCGCTGCTCTCCAGACTTTACAGCAGATGCGTGAT 1339

RESULT 11  
US-10-156-660-3  
; Sequence 3, Application US/10156660  
; Publication No. US20030103958A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay M.  
; APPLICANT: Kretz, Keith  
; APPLICANT: Gray, Kevin A.  
; APPLICANT: Barton, Nelson R.



```

; APPLICANT: Garrett, James B.
; APPLICANT: O'Donoghue, Eileen
; APPLICANT: Mathur, Eric J.
; TITLE OF INVENTION: PHRYASES, NUCLEIC ACIDS ENCODING THEM
; TITLE OF INVENTION: PHRYASES, NUCLEIC ACIDS ENCODING THEM
; FILE REFERENCE: 09010-029007
; CURRENT APPLICATION NUMBER: US/10/156.660
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: US 09/866,379
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 09/580,515
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: US 09/259,214
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: US 08/910,798
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1901
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (188)...(1483)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 403
; OTHER INFORMATION: n = A,T,C or G
US-10-156-660-3

```

Query Match	17.0%; Score 225.2; DB 15; Length 1901;
Best Local Similarity	53.2%; Pred. No. 1.6e-62;
Matches	571; Conservative 0; Mismatches 489; Indels 13; Gaps 4;
QY	107 TAGAACGTGTGGTTATTTTGGATGCCATGTGTCTCGCGCGACCAA-ACAAACGACGC 165
DB	276 TCGAAGTGTGTGATTTGTCTAGTGTGTGCTCTCAACCAAGCCACGCACAC 335
QY	166 TTATGAATGATGTTACGCCACAGATAAGTGGCCACAATGGCCGCTAAAGCGGGTATTAA 225
DB	336 TGTATGCAGGATGTCAACCCACAGCATGGCCAACTGGCCGGTAAAACTGGGTGGCTGA 395
QY	226 CGCCACGTGTGGGAGTGTGTCAATTGATGGGGGGTTTTATGTGTGATTACTTTCGA 285
DB	396 CACCGCGNGTGGTGAGCTAATCGCCCTATCTCGACANTTACCAACGCCACGCTGTGGTAG 455
QY	286 GCGTCTGGTTTTGTTACGC---GCGGGATGTCTCGGCAGAGGGGGGGTATATGCACAGGCAG 342
DB	456 CCGACGGATTGCTGSCGHAANAAGGCTGCCCGCAGTCTGTCAGGTTCGCGATTATTGCTG 515
QY	343 ATATCGATCAAGTACCGCTTTACCGGAACGGCATTTCTTTGATGTGTGGCTTCGGGGT 402
DB	516 ATGTGACAGAGGTACCCGCTAAACAGCGCAAGCCTTCGCGCGCGGGCTGGCACCTGACT 575
QY	403 GTGGTTTGACCGTGCATTAATCAGGCGGATTTGAAAAGACCGATCCCTGTTTCCATCCGG 462
DB	576 GTGCAATAACCGTACATACCCAGGCAGTACGTCCAGTCCCGATTCGTTATTATTATCTCT 635
QY	463 TAGACATGGCGTGTGTAAAGTTAGACAAACGCAACACAGATAAAGCGATTTGAAGAACGAT 522
DB	636 TAAAAACTGGCCTTTTGCCAACTGGATAAACGGAACGTGACTGACGCGATCCCTCAGCAGGG 695
QY	523 TGGCGGGCGGTTAGATACGCTGAGCCACGCGCTACGCTAAACCTTTTGGCCAGATGGGGG 582
DB	696 CAGGAGGGTCAATTGCTGACTTTTACCGGGCATTCGGCAACCGGCGTTTCGCGAAGCTGGAAC 755
QY	583 ACGTGTGTAATTTTCGGCTTCTCCTTTATTGCAAACTTTTGCACACGACGAAGAAAAAAGCT 642

```

RESULT 12
US-10-282-122A-7167
US-10-282-122A, Application US/10282122A
Sequence 7167, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Chilsen, Karl
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELIITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-03-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625

```

US-10-021-723D-3.info

PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 7167  
LENGTH: 1299  
TYPE: DNA  
ORGANISM: Escherichia coli  
US-10-282-122A-7167

Query Match  
Best Local Similarity 53.2%; Pred. No. 2e-62;  
Matches 571; Conservative 0; Mismatches 489; Indels 13; Gaps 4;

QY 107 TAGAAGTGTGGTATTTCAGTCGCCATGCTGCTCGCCGACCA-ACAAGCGAGC 165  
DB 89 TGAAGAAGTGTGGTATTTCAGTCGCCATGCTGCTCGCCGACCA-ACAAGCGAGC 148  
QY 166 TTATGAATGATGTACGCCAGATAGTGGCCACAATGGCCGGTAAAGCGGGGTATTTAA 225  
DB 149 TGATGAGGATGTACGCCAGATAGTGGCCACAATGGCCGGTAAAGCGGGGTATTTAA 208  
QY 226 CGCACGTGTGGGAGTGTGTCACATGATGGGGGGTATTATGCTGATTACTTTGCGA 285  
DB 209 CACCGCGGTGTGAGCTAATCGCTATCTCGACATTAACCAACCGCCGCTGCTGGTAG 268  
QY 286 GCCTTGCTTTGTTAGCG--GCGGGATGTCGCCGAGAGGGGTGCTGCTGCGGATTTGCTG 342  
DB 269 CCGACGATTTGCTGCGAAGAGGGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 328  
QY 343 ATATCGATCAAGTACCGCTTTACCGGACAGAGATTTCTTGATGCTGCTGCTGCTGCTG 402  
DB 329 ATGTCGACGAGCTACCGCTTTAAACAGAGCGGATTCGCGCGGGTGGCACTGACT 388  
QY 403 GTGCTTTGACCGTCAATACAGCGGATTTGAAAGAGCGATCCCTGTTCCATCCG 462  
DB 389 GTGCAATACCGTACATACCGGACAGATAGTCCAGTCCGATCGTATTATTAATCTC 448  
QY 463 TAGAGACTGCGGTGTGAATTAGACCAACCGCAACAGATTAAGGAGTGAAGACGAT 522  
DB 449 TAAAACTGGGTTTGGCAACTGATTAACGCAACGCTGACTGACGGGATCCTCAGCAGG 508  
QY 523 TGGCGGGCGTTAGATACGCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCTG 582  
DB 509 CAGGAGGCTCAATGCTGACTTTACCGGGCATCGGCAACCGGCTTTCCGGAAGTGAAC 568  
QY 583 ACCTGCTCAATTTTGGCGCTTCTCTTTATTCGAAATCTTTGCAACGAGGAGGAGAAACGT 642  
DB 569 GGGTGTCTAATTTTCCGCAATCAACTTTGCTTAAACGCTGAGAAACAGGACGAAAGCT 628  
QY 643 GTGATTTTGGCCATTTTGGCGCCAACTGAAGTTAAGTTAATTAAGAGGAGGACAAAGTGA 702  
DB 629 GTTCATTAACGAGGCAATTAACATCGGAATCAAGGTTGAGCGCGA-----CAATGTCT 682  
QY 703 CCCTCAGTGGGCCACTGGGCTTATCATCGACATTTGGGTGAATCTTCTTATTACAAAACG 762  
DB 683 CATTAACCGGTGCGGTGAAGCTTCGATCAATGCTGACGAGATATTCTCTCTCAACAG 742  
QY 763 CAAAGCCATCGCAGAGGTTCCCTGGCAACGCTTAAAGGGGGGAGGAAATGGGTATCTT 822  
DB 743 CACAGGGAATTTCCCGGAGCGGGGTGGGGAAGGATCACCGATTACACAGTGGAAACACT 802  
QY 823 TATTGTCAATACAAACGCGCAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTT 882  
DB 803 TGCTAAGTTTGCATTAACCGGCAATTTTATTTGCTTACAAACGCAACGCAAGAGGTTGCCG 862  
QY 883 ATAAAGGAGCGCCATTATTACAGCAGATAGATACGGCTTTAAACCTTTCAACCTGATGCTC 942

DB 863 GCOCGCCACCCCGTTATTAGATTTCAGATCAAGACAGCTTGACGCCCATC---CACCGC 919  
QY 943 AGGGGCAAAAGCTACCAATTTAGCCCAAAACCGGGTCTTTTCTCGTGGGATGATA 1002  
DB 920 AAAAACAGGGGTATGCTGCTGACATTAACCACTTCAGTGTCTTTATCGCGGACACGATA 979  
QY 1003 CCATATTTCCCAATATTGCGGGTATGCTGGAGCCGCTGCGAGCTACCGGAGCAACTG 1062  
DB 980 CTAATCTGGCAATCTCGCGCGCACTGGAGCTCAACTGGACGCTTCCCGGTCAGCGG 1039  
QY 1063 ATAATACTCCACAGGTGGGGGATTTGTTTGAACCTATGCAAAATCCGGATTAACACC 1122  
DB 1040 ATAACACGCCGCAAGTGTGTAACCTGCTGTTGAACGCTGGCGTGGCTAAGCGATAACA 1099  
QY 1123 AGCCTAGCTGCGGTGAAGATGTTTACCAAAAGATGATGATGCTGCTGCTGCTGCTG 1175  
DB 1100 GCCAGTGAATTCAGGTTTCGCTGCTTCCAGACTTTTACAGCAGATGCGGTGAT 1152

RESULT 13  
US-09-866-379-9  
Sequence 9, Application US/09866379  
Patent No. US20020136754A1  
GENERAL INFORMATION:  
APPLICANT: SHORT, Jay  
APPLICANT: KRETZ, Keith  
APPLICANT: GRAY, Kevin  
APPLICANT: BARTON, Nelson  
APPLICANT: GARRETT, James  
APPLICANT: O'DONOGHUE, Eileen  
TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYTASES AND USES THEREOF  
FILE REFERENCE: DIVER1370-7  
CURRENT APPLICATION NUMBER: US/09/866,379  
CURRENT FILING DATE: 2001-05-24  
PRIOR APPLICATION NUMBER: US 09/580,515  
PRIOR FILING DATE: 2000-05-25  
PRIOR APPLICATION NUMBER: US 09/318,528  
PRIOR FILING DATE: 1999-05-25  
PRIOR APPLICATION NUMBER: US 09/291,931  
PRIOR FILING DATE: 1999-04-13  
PRIOR APPLICATION NUMBER: US 09/259,214  
PRIOR FILING DATE: 1999-03-01  
PRIOR APPLICATION NUMBER: US 08/910,798  
PRIOR FILING DATE: 1997-08-13  
NUMBER OF SEQ ID NOS: 10  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 9  
LENGTH: 1901  
TYPE: DNA  
ORGANISM: Escherichia coli  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)..(1901)  
OTHER INFORMATION: n is any nucleotide  
US-09-866-379-9

Query Match  
Best Local Similarity 16.9%; Score 223.6; DB 9; Length 1901;  
Best Local Similarity 53.1%; Pred. No. 5.4e-62;  
Matches 570; Conservative 0; Mismatches 490; Indels 13; Gaps 4;

QY 107 TAGAAGTGTGGTATTTCAGTCGCCATGCTGCTCGCCGACCA-ACAAGCGAGC 165  
DB 276 TGAAGAAGTGTGGTATTTCAGTCGCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 335  
QY 166 TTATGAATGATGTACGCCAGATAGTGGCCACAATGGCCGGTAAAGCGGGGTATTTAA 225  
DB 336 TGATGAGGATGTACGCCAGATAGTGGCCACAATGGCCGGTAAAGCGGGGTATTTAA 395  
QY 226 CGCACGTGTGGGAGTGTGTCACATGATGGGGGGTTCATGCTGCTGCTGCTGCTGCTGCTG 285  
DB 396 CACCGCGGGTGTGAGCTAATCGCTATCTCGGACATTAACCAACGCGGCTGCTGCTG 455

QY 286 GCCTTGGTTGTTAGCG---CCGGGATGTCGGCAGAGGGGGGGTATATGACAGGCAG 342  
Db 456 CCGACGGAATTCCTCGCGGAAAAGGCTGCGCGAGTCTGTGTCAGTCGCGGATTTGCTG 515  
QY 343 ATATCGATCAACGTAACCGCTTAAACCGACAGGCAATTTCTTGATGTTGCTGCGGCTG 402  
Db 516 ATGTCGACGAGCGTACCGTAAACAGCGGAGCGCTTTCGCGCGGCTGCGACCTGACT 575  
QY 403 GTGGTTGACCGTGCATTAATCAGCGCGATTTGAAAAGACCGATCCCTCTTCCATCGG 462  
Db 576 GTGCAATAACCGTACATACCCAGGAGATAGCTCCAGTCCGATCCGTTATTTAATCCT 635  
QY 463 TAGAGACTGGCGTGTGAAGTTAGACACGACAGCAACAGATAAAGCGATTTGAAGAAGCAT 522  
Db 636 TAAAAACTGGCTTTGCCAATCGATTAACGCGAAGCGTGAAGCGATCTCAGCAGGG 695  
QY 523 TGGGGGGCGCTTAGATAGCGTAGCGAGCGCTAGCTAAACCTTTTGGCCAGATGGGG 582  
Db 696 CAGGAGGCTCAATTGCTGACTTTACCGGGCATCGGCAACGGCGTTTTCGGAACTGAAC 755  
QY 583 ACGTGTCAATTTTGGCGCTTCTCGTTATTCGCAATCTTTGCAACAGCAGGAAAACGT 642  
Db 756 GGGTGTCTTAATTTTCGCAATCAAACTTGTCCTTAAACGTGAGAAACAGGACGAACT 815  
QY 643 GTGATTTTGGCCACTTTTCGGGCAATGAAGTTAAGTTAATAAAGAGGAGCAAAAGTGA 702  
Db 816 GTTCAATTAACGAGGCAATTAACCATCGGAATCAAGGTGAGCGCGA-----CAATGCT 869  
QY 703 CCTCAGTGGCGCACTGCGCTTATCATCGACATTTGGGTGAATCTTCTTATTACAAACG 762  
Db 870 CATTAACCGTGGGTAGCTTCGATCATGCTGACGAGATATTTCTCTGCAACAG 929  
QY 763 CACAAGCCATGCGAGGTTGCTGGCAACGGCTAAAGGGGGGAGAAATGGGTATCCT 822  
Db 930 CACAGGGAATGCGGAGCGGGGTGGGAAGGATCACCGATTACACCGTGAACACCT 989  
QY 823 TATTCGTATTAATCAACGCGCAATTTAATTTAATGGCAAAACACCGTATATCCCGCTC 882  
Db 990 TCTAAGTTGCAATTAACGCGCAATTTGTTGCTACACGACGCGCAGAGTTGCGCGCA 1049  
QY 883 ATAAAGGAGCGCAATTTATACAGCAGATAGATAGCGCTTTAAACCTTCAACTGATGCT 942  
Db 1050 GCGCGGCAACCGCTTATAGATTTGATCAAGACAGCGCTTACGCGCCCATC---CACCG 1106  
QY 943 AGGGGCAAAAGCTACCAATTTACGCGCAACCGGCTTCTGCTCGGTGGGATGATA 1002  
Db 1107 AAAAAACAGCGGTATGGTGTGACATTACCCACTTTCAGTGTGTTATTCGCCGACAGATA 1166  
QY 1003 CCAATATTGCCAATATTGGGGTATGCTGGGAGCGGCTGCGAGCTACCCGAGCAACCTG 1062  
Db 1167 CTAACTGCAAAATCTCGCGCGGCACTGGAGCTCAACTGGACGCTTCCCGGTACGCGG 1226  
QY 1063 ATAATACTCACAGGTGGGGAATGGTTTGAACCTATGCAAAATCGGATACCAACC 1122  
Db 1227 ATAACAGCGCGGAGGTGGTGAAGTGGTTTGAACGCTGGGCTGCGGCTAAGCGGATAACA 1286  
QY 1123 AGCGCTACGTTGCGGTGAAGATGTTCTACCAACAGATGGATCAGTTGGGTAA 1175  
Db 1287 GCCAGTGGATTGAGTTTCGCTGGCTTCCAGACTTACAGCAGATGGGTAT 1339

RESULT 14  
US-09-777-566A-1  
; Sequence 1, Application US/09777566A  
; Patent No. US20010055788A1  
; GENERAL INFORMATION:  
; APPLICANT: DIVERSA CORPORATION  
; APPLICANT: SHORT, Jay  
; APPLICANT: KRETZ, Keith  
; TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYTASES AND USES THEREOF  
; FILE REFERENCE: DIVER1370-6  
; CURRENT APPLICATION NUMBER: US/09/777,566A

; CURRENT FILING DATE: 2001-06-11  
; PRIOR APPLICATION NUMBER: US 09/318,528  
; PRIOR FILING DATE: 1999-05-25  
; PRIOR APPLICATION NUMBER: US 09/291,931  
; PRIOR FILING DATE: 1999-04-13  
; PRIOR APPLICATION NUMBER: US 09/259,214  
; PRIOR FILING DATE: 1999-03-01  
; PRIOR APPLICATION NUMBER: US 08/910,798  
; PRIOR FILING DATE: 1997-08-13  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 1  
; LENGTH: 1323  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1320)  
; NAME/KEY: misc.feature  
; LOCATION: (1)..(1323)  
; OTHER INFORMATION: n is any nucleotide  
US-09-777-566A-1  
  
Query Match 16.5%; Score 218.8; DB 9; Length 1323;  
Best Local Similarity 52.8%; Pred. No. 1.6e-60;  
Matches 567; Conservative 0; Mismatches 493; Indels 13; Gaps 4;  
  
QY 107 TAGAAGCTGTGTTATTTAGTCGCCATGTCGCTGCGCGACCAA-ACAAGCGAGC 165  
Db 89 TGAAGAGTGTGCTGATTTGTCAGTCGTCATGTCGTCCTCAACCAAGGCGCACCAAC 148  
QY 166 TTATGATCATGTTAGTCGCCAGATAGTGGCCCAATAGTGGCGTAAAGCGGGTATTTAA 225  
Db 149 TGATGAGATGTCACCCGACAGCATGSCCACTGSCCGGTAAACTGGTGGCTCA 208  
QY 226 CCGCACGTGTGGGAGTTGGTCACATGATGGGGGGTTTATGATGGTGTGCTCGGGGT 285  
Db 209 CACCGGCGTGTGAGCTAATCGCTATCTCGGACATTTACCAAGCCGACGCTCTGGTAG 268  
QY 286 GCCTTGGTTGTTAGCG---GCGGGATGTCGCGAGAGGGGGGTATATGACAGGCAG 342  
Db 269 CCGACGATTTGCGGAAAAAGGGCTGCGCGAGTCTGTCAGTCTGCGCATTTTGTCTG 328  
QY 343 ATATCGATCAACGTAACCGCTTTAAACCGGACAGGCATTTCTTGATGGTGTGCTCGGGGT 402  
Db 329 ATGTGACGAGCGTACCGTAAACAGGGGAGCGCTTCGCCGCGGGCTTGGCACTGACT 388  
QY 403 GTGGTTGACCGTGCATTAATCAGCGCGATTTGAAAAGACCGATCCCTGTTCATCCCG 462  
Db 389 GTCAATAACCGTACATACCCAGGAGATACGTCAGTCCGATCCGATCGTTATTTAATCCTC 448  
QY 463 TAGAGACTGCGGTGTGAAGTTAGACAACGCAACAGATAAAGCGATTGAAGACGAT 522  
Db 449 TAAAACTGGCGTTTGCCAACTGGATAACGCGAAGCGTACTGACGCGATCCTCAGCAGG 508  
QY 523 TGGGGCGCGTTAGATAGCGTAGCGGCTACGCTAAACCTTTTGGCCAGATGGGG 582  
Db 509 CAGGAGGGTCAATTTGCTGACTTTACCGGGCATCGGCAACCGGCTTTCGGAACCTGGAAC 568  
QY 583 ACGTGTGATTTTGGGGCTTCTCTTATTTGCAAAATCTTTTGCAACAGCAAGGAAAAACGT 642  
Db 569 GGGTGTCTAATTTTCGCAATCAAACTTGTGCTTAAACGTGAGAAACAGGACGAAAGCT 628  
QY 643 GTGATTTGCGCACTTTGGCGGCAATGAAGTTAAAGTTAAAGAGGACGCAAAAGTGA 702  
Db 629 GTTCAATTAACGAGGCAATTAACATCGGAATCAAGGTGAGCGCGA-----CAATGTCT 682  
QY 703 CCCTCAGTGGGCGCACTGGGGTTATCATCGACATTTGGGTGAAATCTTTTATTACAAAACG 762  
Db 683 CATTAACCGGTGGGTAAAGCTCGCATCAATGCTGAGGAGATATTTCTCTGCAACAG 742  
QY 763 CACAAGCCATGCCAGGTTGCTCGCAACGGCTTAAAGGGGCGGAGAAATGGGTATCCT 822



Job time : 572.19 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2004, 22:18:08 ; Search time 2501.04 Seconds  
(without alignments)  
15820.390 Million cell updates/sec

Title: US-10-021-723B-3  
Perfect score: 1325  
Sequence: 1 agctgatttagataatcg.....aacctgcctgcacatctaa 1325

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues  
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pin:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_pbg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	115.2	8.7	529	13	BU713770
c	2	107.4	8.1	616	12
	3	103.4	7.8	746	13
c	4	90.6	6.8	595	14
					BU713770 SJAABUG02
					BJ074127 BJ074127
					CA093060 SCCOCL200
					CF326092 JMT11 -05-

C	5	75.4	5.7	393	10	BE520240
	6	63.4	4.8	531	13	BU713769
	7	62.6	4.7	383	9	AK036132
c	8	62.6	4.7	383	9	AK036134
	9	50.4	3.8	885	13	EX425603
	10	49.2	3.7	707	12	BU619443
c	11	48.8	3.7	208	12	BM409016
c	12	48.8	3.7	208	12	BM412806
c	13	46.2	3.5	354	9	AA545747
c	14	44	3.3	1201	13	BM381961
c	15	43.2	3.3	1101	29	CNS00KX2
c	16	41.8	3.2	520	13	EX477052
c	17	40.8	3.1	858	13	BU930958
c	18	40.4	3.0	179	10	BE461872
c	19	40.2	3.0	553	9	AU280373
c	20	40.2	3.0	594	13	EX473143
c	21	40.2	3.0	602	13	EX479427
c	22	40.2	3.0	626	9	AL701247
c	23	40.2	3.0	643	10	BE314719
c	24	40.2	3.0	649	10	BE313446
c	25	40.2	3.0	688	10	BE298415
c	26	40.2	3.0	692	13	EX482609
c	27	40.2	3.0	722	9	AL709367
c	28	40.2	3.0	893	13	BU931710
c	29	40.2	3.0	896	10	BE796373
c	30	40.2	3.0	899	13	EX386580
c	31	40.2	3.0	954	13	BU845888
c	32	40.2	3.0	969	13	EX333117
c	33	40.2	3.0	989	10	BE314590
c	34	40.2	3.0	1064	12	BM912235
c	35	40.2	3.0	1067	13	EX397055
c	36	40.2	3.0	1126	12	BM462187
c	37	40.2	3.0	1139	12	BM916904
c	38	40.2	3.0	1171	13	EX394513
c	39	40.2	3.0	1201	9	AL529590
c	40	40.2	3.0	1201	13	EX354373
c	41	40.2	3.0	1705	12	BM456811
c	42	39.8	3.0	1201	13	EX385091
c	43	39.6	3.0	1190	13	EX377026
c	44	39.4	3.0	723	13	EX727085
c	45	39.4	3.0	992	13	EX394359

## ALIGNMENTS

RESULT 1  
BU713770  
LOCUS  
DEFINITION

BU713770 529 bp mRNA linear EST 23-OCT-2003  
SJAABUG02 Adult sJc 7/94 Schistosoma japonicum cDNA similar to  
pdb|1DKPIA Chain A, Crystal Structure Of Phytate Complex Of  
Escherichia Coli Phytase At Ph 6.6. Phytate Is Bound With Its  
3-Phosphate In The Active Site. Hg2+ Cation Acts As An  
Intermolecular Bridge, mRNA sequence.

ACCESSION BU713770 GI:28321126  
VERSION BU713770.1  
KEYWORDS EST.  
SOURCE Schistosoma japonicum  
ORGANISM Schistosoma japonicum

REFERENCE  
AUTHORS

Hu, W., Yan, Q., Shen, D. K., Liu, F., Zhu, Z. D., Song, H. D., Xu, X. R., Wang, Z. J., Rong, Y. P., Zeng, L. C., Wu, J., Zhang, X., Wang, J. J., Xu, X. N., Wang, S. Y., Fu, G., Zhang, X. L., Wang, Z. Q., Brindley, P. J., McManus, D. P., Xue, C. L., Feng, Z., Chen, Z. and Han, Z. G.  
Evolutionary and biomedical implications of a Schistosoma japonicum complementary DNA resource  
Nat. genet. 35 (2), 139-147 (2003)

TITLE  
JOURNAL  
MEDLINE  
COMMENT

Contact: Zeguang Han  
Chinese National Human Genome Center at Shanghai  
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai

201203, P. R. China  
Tel: 86-21-50801919 (ex. 45)  
Fax: 86-21-50801922  
Email: hanzg@chc.sh.cn.

## FEATURES

source

1. 529  
/organism="Schistosoma japonicum"  
/mol\_type="mRNA"  
/strain="Chinese (Anhui) strain"  
/db\_xref="taxon:6182"  
/sex="Male and female"  
/tissue\_type="Whole body"  
/dev\_stage="Adult worms"  
/lab\_host="Mouse and rabbit"  
/clone\_lib="Adult SJC 7/94"  
/note="Vector: Lambda ZAP-II XR.; Site 1: EcoR I; Site 2:  
XhoI I; Several hundred adult Schistosoma japonicum  
(Anhui, P.R. China, strain), of mixed sex, were perfused  
from the mesenteries of experimentally infected mice and  
rabbits at the Queensland Institute of Medical Research,  
Brisbane, Australia (QIMR), and stored for several months  
in liquid nitrogen. Subsequently, mRNA was isolated at the  
QIMR from lysates of these worms by oligo dT  
chromatography, using a kit from Pharmacia. The mRNA was  
then shipped to Clontech, Palo Alto, CA, USA, who  
constructed a cDNA library. First strand synthesis was  
primed with an oligo-dT-XhoI primer and synthesized using  
M-MuV reverse transcriptase. Second strand synthesis was  
accomplished with RNase H and T4 DNA polymerase. The  
double stranded cDNA was ligated to EcoRI linkers,  
digested with EcoRI and XhoI, and ligated into the  
phagemid vector lambda ZAP II XR. After construction of  
this directional library by Clontech, it was returned to  
the QIMR. During analysis of the library at the QIMR, we  
have found that a small percentage, 2% to 3%, of the  
clones contain inserts that appear to be highly homologous  
to sequences from salmonid fishes, as determined by  
homology comparisons using BLAST and by Southern  
hybridization analysis to genomic DNA from salmon (Sigma  
Chemical Co., St. Louis, MO) under stringent washing  
conditions. The remainder of the clones appear to contain  
S. japonicum sequences."

## ORIGIN

Query Match 8.7%; Score 115.2; DB 13; Length 529;  
Best Local Similarity 54.0%; Pred. No. 5.1e-25;  
Matches 259; Conservative 0; Mismatches 218; Indels 3; Gaps 1;  
QY 696 AAAGTGACCCCTCAGTGGGCCACTGGCGTTATCATCGACATGGGTGAATCTTCTTATTA 755  
DB 7 AATGTCCTATTAAACCGGTGGGTAAAGCGCTGCATCAATGCTGACGGAGATATTTCTCTG 66  
QY 756 CAABAAGCACAAGCCATCCAGAGTTTCCTGGCAACGGCTAAAGGGGGGAGAAATGG 815  
DB 67 CAACAAGCACAGGGAATCCGGAGCCGGGGTGGGGAAGGATCAACCGATTACACCACTGG 126  
QY 816 GTATCCTTTATGTCTATTACATAAAGCGCAATTTAATTTAATGGCAAAAACACCGTATATC 875  
DB 127 AACACCTTGTAAAGTTCATACCGGCAATTTATTTGTCTACACGACGCGAGAGTT 186  
QY 876 GCCCGTCATAAAGGAGCGCAATTTATTAAGCAGATAGATACGGCTTTAAACCTTCAACTG 935  
DB 187 GCCCGCAGCGCGCCACCGCTTATTAATTTGATCAAGACAGGGTTGACGCCCCCATC-- 244  
QY 936 GATGCTCAGGGGCAAAAGCTACCCATTTTCAGCCCAAAACCGGGTCTTTCTCTGGTGGG 995  
DB 245 -CACCGCAAAACAGCGGTATGGTGTGACATTTACCATCTCAGTGTCTTTATCCCGGA 303  
QY 986 CATGATACCAATATTGCGCAATATTGCGGATATGCTGGAGCGGACTGCGAGTACCGGAG 1055  
DB 304 CACGATATAATCTGGCAATCTCGCGCGCACTGGAGCTCAACTGGAGCTTCCCGGT 363  
QY 1056 CBACTGTATACTCCACGAGTGGGGATGGTTTTTGTAGTACTGCAAAATCCGGAT 1115

DB 364 CAGCCGGATATACAGCCGCCAGGTGGTGAATGTTTGAACGCTGGCGTGAAGC 423  
QY 1116 AACCAACAGCCGCTACGTTTGGGTGAAGATGTTTACCAAAACGATGGATCTGGTGAAT 1175  
DB 424 GATAACAGCCAGTGGATTCAGGTTTGGCTGCTTCCAGACTTTTACAGCAGATGGTGAT 483  
RESULT 2  
BJ074127/c 616 bp mRNA linear EST 29-SEP-2003  
LOCUS BJ074127 NIBB Mochii normalized Xenopus tailbud library Xenopus  
DEFINITION laevis cDNA clone XL090j06 5', mRNA sequence.  
ACCESSION BJ074127  
VERSION BJ074127.1 GI:17504316  
KEYWORDS EST.  
SOURCE Xenopus laevis (African clawed frog)  
ORGANISM Xenopus laevis  
REFERENCE 1 (bases 1 to 616)  
AUTHORS Kitayama, A., Terasaka, C., Mochii, M., Ueno, N., Shin-i, T. and  
Kohara, Y.  
TITILE Expressed genes in X. laevis embryo  
JOURNAL Unpublished (2001)  
COMMENT Contact: Tadasu Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tsunigenes.nig.ac.jp  
The information of this clone is available through the following  
URL.  
http://xenopus.nibb.ac.jp.

## FEATURES

source

1. 616  
Location/Qualifiers  
/organism="Xenopus laevis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8355"  
/clone="XL090j06"  
/tissue\_type="whole embryo"  
/dev\_stage="stage 25"  
/clone\_lib="NIBB Mochii normalized Xenopus tailbud  
library"

## ORIGIN

Query Match 8.1%; Score 107.4; DB 12; Length 616;  
Best Local Similarity 50.1%; Pred. No. 1.8e-22;  
Matches 298; Conservative 0; Mismatches 291; Indels 6; Gaps 1;  
QY 305 GGGATGTCGGCAGAGGGGGGTATATGCACAGCAGATATCGATCAACGTACCGCTT 364  
DB 589 GGGGTGCCCCGAGTCTGGTCAGGTCGGATTTATTCGATGTCGACGAGCGTACCCTGTA 530  
QY 365 AACCGCACAGGCATTTCTTTGATGGTGTGGCTCCGGGGTGTGGTTTGACCGTCATAATCA 424  
DB 529 AACGGCGAAGCTTCGCGCGCGGCTGGCACCCTGACTGTGCAATAACCGTACATACCCA 470  
QY 425 GGCGGATTTGAAAAGACCGATCCCTGTTCCATCCGTAGAGACTGGGTTGTAGTT 484  
DB 469 GGCAGATACGTCAGTCCCGATCCGTTTAACTCTCTAAAACTGGGTTTGGCACT 410  
QY 485 AGACACGCAACAAACAGATATAAGCGATTGAAGAACGATTGGCGGGCCGTTAGATACGGT 544  
DB 409 GGATAACGCGAAGCTGACTGACGCGATCCTCAGCAGGGCAGGAGGTCAATTGTGACTT 350  
QY 545 GAGCAGCGCTACGGCTAAACCTTTTGGCCAGATGGGGACGTCGTGTAATTTTGGCGCTTC 604  
DB 349 TACCGGCACTCGGCAAAACGGCGTTTCGCGAAGTGGAAACGGGTGCTTAATTTTCCCAATC 290  
QY 605 TCCTTATTGCAATCTTTTGCACAGCAAGAAAAACGTGTGATTTTGCACCACTTTGGCGSC 664



```
Db      289  AAACTTGTGCTTAAACGTCGAGAAACAGGACGAAAGCTGTTCAATTAACGAGGCAATTACC 230
QY      665  CAATGAAGTTAAAGTTAATAAGAGGACAAAGTGACCTCAGTGGGCGCACTGGCGTT 724
Db      229  ATCGGAACTCAAGTGTAGCCCGA-----CAATGCTCAATTAACCGGTGCGGTAGCGCT 176
QY      725  ATCATGACATTTGGGTGAAATCTTCTTATTACAAAACGCAAGCCATCCAGAGTTGC 784
Db      175  CGCATCAATCTGACGGAGATAATTTCTCTGCAACAAAGCACTGGGAATGCCGAGCCGGG 116
QY      785  CTGCGACGCTAAAGGGGCGGAGATGGGTATCTTATTGCTCATTAATCATTAACGCGCA 844
Db      115  GTGGGAGGATCACCGATTACACCACTGTGAACACCTTGCTAAGTTTGCATTAACCGCA 56
QY      845  ATTTAATTTTATGGCAAAAACACCGTATATCGCCCGTCATAAAGGACGCCATT 899
Db      55  ATTTTATTGCTACAACGCGCAGAGGTTGCCCGCAGCGCGCCACCCCGTTA 1
```

```
RESULT 3
CA093060      746 bp mRNA linear EST 23-SEP-2003
LOCUS      SCCCL2001D10.b CL2 Saccharum officinarum cDNA clone SCCCL2001D10
DEFINITION      3', mRNA sequence.
```

```
ACCESSION      CA093060.1 GI:34946367
VERSION
KEYWORDS
SOURCE
ORGANISM      Saccharum officinarum
              Saccharum officinarum
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
              clade; Panicoideae; Andropogoneae; Saccharum.
```

```
REFERENCE
AUTHORS      Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
TITLE      The libraries that made SUCEST
JOURNAL      Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
COMMENT      Contact: Arruda P
              Centro de Biologia Molecular e Engenhariaia Genetica
              Universidade Estadual de Campinas
              Caixa Postal 6010, 13083-970, Campinas SP, Brazil
              Tel: 55 19 3788 1137
              Fax: 55 19 3788 1089
              Email: parruda@unicamp.br
              Clone distribution: clone distribution information can be found
              through the Brazilian Clone Collection Center (BCCC) at
              http://www.bcccenter.fcav.unesp.br
              Plate: 001 row,D column: 10
              Seq primer: M13/Forward primer.
```

```
FEATURES
source
1..746
Location/Qualifiers
    /organism="Saccharum officinarum"
    /mol_type="mRNA"
    /db_xref="taxon:4547"
    /clone="SCCCL2001D10"
    /lab_host="XlBlue MRF"
    /clone_lib="CL2"
    /note="Organ: Pool of sugarcane calli submitted to low
    (40C) and high (37 C) temperature stress; Vector:
    pBluscript; Site 1: EcoRI; Site 2: XhoI; An
    unidirectional cDNA library generated from [Pool of
    sugarcane calli submitted to low (40C) and high (37 C)
    temperature stress]. cDNA was prepared from polyA+ mRNA
    using ZAP - cDNA Synthesis Kit (Stratagene). The
    double-strand cDNAs were fractionated in a sepharose CL-2B
    40cm-columns and fragments sizing between 0.8 and 1.5 Kb
    were directionally cloned into the vector. Details of
    each source of RNA and library construction can be
    obtained at http://sucest.lad.ic.unicamp.br/public"
```

```
ORIGIN
Query Match      7.8%; Score 103.4; DB 13; Length 746;
Best Local Similarity 53.5%; Pred. No. 3.8e-21;
```

```
Matches 257; Conservative 0; Mismatches 219; Indels 4; Gaps 2;
QY      696  AAAGTGACCCCTCAGTGGGCGCACTGGCGTTATCATCGCAATTTGGGTGAAATCTTCTTATTA 755
Db      265  AATGCTCATTAACCGGTGCGGTAAAGCTCGCATCAATGCTGACGGAGATATTTCTCTG 324
QY      756  CAAAACGACAAAGCCATGCGAGAGTTCCCTGGCAACGGCTAAAGGGGCGGGAATGG 815
Db      325  CAACAAGCACAGGAATGCGGAGCCCGGGTGGGGAAGGATCACCGATTACACACAGTGG 384
QY      816  GTATCTTATTGTCAATACATAAGCGCAATTTAATTTAATGGCAAAAACCCGTATATC 875
Db      385  AACACCTTGCTAAGTTTGCATAACGCGCAATTTTATTGCTACACGCGCAGAGGTT 444
QY      876  GCGCGTATAAAGGAGCGCCATTTATACAGAGATAGATACGGCTTTAACCCCTTCAACTG 935
Db      445  GCGCGCAGCGCGCCACCCCGTTATTAGATTGATCAAGACAGCGTTGACGCCCATC-- 502
QY      936  GATGCTCAGGCGCAAAAGCTACCCATTTCAGCCAAAACCGGGTCTTGTTCCTCGGTGGG 995
Db      503  -CACCGCAAAACAGCGGTATGGTGTGACATCCCATTCAGTGTCTTTATCGCGGA 561
QY      996  CATGATACCAATATTGCCAATATTGCGGGTATGCTGGAGCGGACTGCGAGCTACCCGAG 1055
Db      562  CACGATATACTATGCGCAATCTCGGGCGGCACTGGAGCTCAACTGGACGCTTCCCGGT 621
QY      1056  CAACCTGATATACCTCCACAGGTGGGGATTGTTTGAACATATGCAAAATCCGGAT 1115
Db      622  CAGCCGATAACACGCCCGCAGGTGGTGAATGG-TGTTGAACGCTGGCGTNGCTAAGC 680
QY      1116  AACCACCAAGCGCTAGTTGCGGTGAAGATGTTTACCAAAACGATGGATCAGTTGGTAAAT 1175
Db      681  GATAACACCAAGTGGATTCAGGTTTCGCTCTCCAGACTTTTACAGCANATGCGTGTAT 740
```

```
RESULT 4
CF326092/c
LOCUS      CF326092
DEFINITION      JMT1--05-B09-g1 AtJMT-overexpressing transgenic rice lambda phage
              cDNA library (JMT1) Oryza sativa cDNA clone JMT1--05-B09, mRNA
              sequence.
```

```
ACCESSION      CF326092
VERSION
KEYWORDS
SOURCE
ORGANISM      Oryza sativa
              Oryza sativa
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
              Ehrhartoideae; Oryzae; Oryza.
              1 (bases 1 to 595)
              Kim,J.S., Jun,K.M., Cheong,P.J., Kim,M.J., Lee,T.H., Shin,Y.C.,
              Song,S.I., Kim,J.K., Kim,Y.-K. and Nahm,B.H.
              Large-scale Sequencing Analysis of Rice ESTs
              Unpublished (2003)
              Contact: Nahm B.H.
```

```
REFERENCE
AUTHORS      Genomics and Genetics Institute, GreenGene Biotech Inc.; Division
              of Bioscience and Bioinformatics, Myongji University
              Yongin, Kyeonggi, Korea
              Tel: 82 31 330 6193
              Fax: 82 31 321 6355
              Email: bhnah@gbio.com, bhnah@bio.myongji.ac.kr.
```

```
FEATURES
source
1..595
Location/Qualifiers
    /organism="Oryza sativa"
    /mol_type="mRNA"
    /cultivar="Nackdong"
    /db_xref="taxon:4530"
    /clone="JMT1--05-B09"
    /tissue_type="leaf"
    /dev_stage="14 days after germination"
    /lab_host="E.coli SOLR"
    /clone_lib="AtJMT-overexpressing transgenic rice lambda
    phage cDNA library (JMT1)"
```

/note="Vector: pBluescript SK(+); Site 1: EcoRI; Site 2: XhoI; cDNA was inserted into lambda Uni-ZAP XR vector at 5' end with EcoRI and 3' end with XhoI site. mRNA was prepared from Arabidopsis Jasminate Carboxyl methyltransferase overexpression line."

## ORIGIN

Query Match 6.8%; Score 90.6; DB 14; Length 595;  
Best Local Similarity 55.9%; Pred. No. 4.3e-17;  
Matches 193; Conservative 0; Mismatches 149; Indels 3; Gaps 1;

QY 831 TTACATAACGCCCAATTAATTAATGCGAATAACACCGTATATGCGCCGTCATAAAGG 890  
Db 592 TTGCATAACGCCCAATTAATTAATGCTACACGACCGCAGAGTTGCGCCGACGCGGCC 533  
QY 891 AGCCATTATTACAGCAGATAGATACGGCTTTAAACCTTTAACTGATGCTCAGGGCAA 950  
Db 532 ACCCGTTATTAGATTGATCAAGACAGCGTTGACGCCCATCCAC---CGCAAAACAG 476  
QY 951 AAGCTACCCATTACGCCCAAAACCGGCTTTGCTTCGCTGGGTCATGATACCAATAT 1010  
Db 475 GCGTAGGTGTGACATTACCCACTTCAGTGTCTGTTTATCGCGGACAGATACATACTG 416  
QY 1011 GCAATATTGCGGCTATGTCGGAGCCGACTGGCAGCTACCGAGCAACCTGATAATACT 1070  
Db 415 GCAATCTCGCGGCGCACTGAGCTCACTGGAGCTTCCGCTCAGCCGATACAG 356  
QY 1071 CCACAGGTGGGGATATGTTTTTGAATATGCGAAAATCGGATACACACAGCGTAC 1130  
Db 355 CGCCAGGTGGTGAATCTGTTTGAACGCTCGGCTAAGCGATAACAGCCAGTGG 296  
QY 1131 GTTGGGTGAAGATGTTCTACCAACGATGATGATGCTGCTAAT 1175  
Db 295 ATTGAGTTTCTGCTGCTTCCAGACTTTACAGCAGATGCGTGAT 251

RESULT 5  
LOCUS B5520240/c 393 bp mRNA linear EST 19-MAR-2001  
DEFINITION M1812STM Arabidopsis developing seed Arabidopsis thaliana cDNA  
clone M1812 5', mRNA sequence.

ACCESSION B5520240.1 GI:9778242  
VERSION B5520240  
KEYWORDS EST.  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 393)  
White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de  
Ilarduya, O., Jaworski, J.G., Ohlrogge, J. and Benning, C.  
A new set of Arabidopsis expressed sequence tags from developing  
seeds. The metabolic pathway from carbohydrates to seed oil  
Plant Physiol. 124 (4), 1582-1594 (2000)

20567808  
11115876  
Contact: Benning, C  
Dept. of Biochemistry & Molecular Biology  
Michigan State University  
224 Biochemistry, Michigan State University, East Lansing, MI  
48824, USA  
Tel: 517 355 1609  
Fax: 517 353 9334  
Email: benning@msu.edu  
Michigan State University DNA Sequencing Facility Arabidopsis  
Biological Resource Center, The Ohio State University, 309 Botany &  
Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:  
6142920603 TEL: 6142929371.

Location/Qualifiers

1. .393  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"

FEATURES  
source

/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone="M1812"  
/tissue\_type="seed"  
/dev\_stage="5-13 days after flowering"  
/lab\_host="E.coli"  
/clone\_lib="Arabidopsis developing seed"  
/notes="Organ: Developing seed; Vector: pBluescript SK-;  
Site\_1: EcoRI; Site\_2: XhoII"

## ORIGIN

Query Match 5.7%; Score 75.4; DB 10; Length 393;  
Best Local Similarity 51.0%; Pred. No. 2.6e-12;  
Matches 175; Conservative 0; Mismatches 168; Indels 0; Gaps 0;

QY 339 GCAGATATCGATCAACGCTACCCGCTTAACGCGACAGCATTTCTTCATCGTGTGGCTCG 398  
Db 387 GCTGATGTCGACGAGCGGTACCCGTAACACAGCGGAGCTTTGCGCCGCGGTTGGCACT 328  
QY 399 GGGTGTGGTTTACCGTCGATATCAGGCGCATTTGAAAAAGACCGATCCCTGTTCCAT 458  
Db 327 GACTGTGCAATACCGTACATACCCAGGCGAGTACGTCAGTCCCGATCCGTTATTAAAT 268  
QY 459 CCGGTAGAGACTGGCGTGTAGTTAGTAAACGCAACAGATTAAGCGATTGAAGAA 518  
Db 267 CCTTAAAACTGGCGTTTCCCACTGATACGGAACGACTGACTGACGCGATCTCAGC 208  
QY 519 CGATTGGCGGCGGCTTAGATACGGTACGCGCTACGCTAAACCTTTGCGCCAGATG 578  
Db 207 AGGCGAGGAGGGTCAATTGCTGACTTTACCGGCGCATCGGCAAAACGGCGTTTCGCACTG 148  
QY 579 GGGGAGCTGCTGAATTTTGGCGTCTCCTTATTGCAATCTTTGCAACAGCAAGGAAA 638  
Db 147 GAACGGGTGCTTAATTTTCCGCAATCAAACTTGTGCTTAAACGTGAGAAACAGGACGA 88  
QY 639 ACGTGTGATTTTGCCTCACTTTGCGGCAATGAAGTTTAACTGTTA 681  
Db 87 AGCTGTTTCATTAACGAGGCAATTAACATCGGAACCTCAGGTGA 45

## RESULT 6

LOCUS BU713769 531 bp mRNA linear EST 23-OCT-2003  
DEFINITION SJABUG01 Adult SJC 7/94 Schistosoma japonicum cDNA similar to  
SP[P07102]PPA ECOLI PERIPLASMIC APPA PROTEIN PRECURSOR (INCLUDES:  
PHOSPHOANHYDRIDE PHOSPHOHYDROLASE (PH 2.5 ACID PHOSPHATASE) (AP);  
6-PHYTASE 1, mRNA sequence.

ACCESSION BU713769  
VERSION BU713769.1 GI:28321125  
KEYWORDS EST.

SOURCE Schistosoma japonicum  
ORGANISM Schistosoma japonicum

Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;  
Strigeidae; Schistosomatoidea; Schistosomatidae; Schistosoma.  
1 (bases 1 to 531)  
Hu, W., Yan, Q., Shen, D.K., Liu, F., Zhu, Z.D., Song, H.D., Xu, X.R.,  
Wang, Z.J., Rong, Y.P., Zeng, L.C., Wu, J., Zhang, X., Wang, J.J.,  
Xu, X.N., Wang, S.Y., Fu, G., Zhang, X.L., Wang, Z.Q., Brindley, P.J.,  
McManus, D.P., Xue, C.L., Peng, Z., Chen, Z. and Han, Z.G.  
Evolutionary and biomedical implications of a Schistosoma japonicum  
complementary DNA resource

Nat. Genet. 35 (2), 139-147 (2003)  
22879925  
Contact: Zeguang Han  
Chinese National Human Genome Center at Shanghai  
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
201203, P. R. China  
Tel: 86-21-50801919 (ex.45)  
Fax: 86-21-50801922  
Email: hanzg@chgc.sh.cn.

JOURNAL MEDLINE  
COMMENT

FEATURES  
source  
1. .531  
/organism="Schistosoma japonicum"

```

/mol_type="mRNA"
/strain="Chinese (Anhui) strain"
/db_xref="taxon:6182"
/sex="Male and female"
/tissue_type="Whole body"
/dev_stages="Adult worms"
/lab_host="Mouse and rabbit"
/clone_lib="Adult Sjc 7/94"
/notes="Vector: Lambda ZAP-II XR.; Site 1: EcoR I; Site 2: XhoI I; Several hundred adult Schistosoma japonicum (Anhui, P.R. China, strain), of mixed sex, were perfused from the mesenteries of experimentally infected mice and rabbits at the Queensland Institute of Medical Research, Brisbane, Australia (QIMR), and stored for several months in liquid nitrogen. Subsequently, mRNA was isolated at the QIMR from lysates of these worms by oligo dT chromatography, using a kit from Pharmacia. The mRNA was then shipped to Clontech, Palo Alto, CA, USA, who constructed a cDNA library. First strand synthesis was primed with an oligo-dT-XhoI-primer and synthesized using M-MuLV reverse transcriptase. Second strand synthesis was accomplished with RNase H and T4 DNA polymerase. The double stranded cDNA was ligated to EcoRI linkers, digested with EcoRI and XhoI, and ligated into the phagemid vector lambda ZAP II XR. After construction of this directional library by Clontech, it was returned to the QIMR. During analysis of the library at the QIMR, we have found that a small percentage, 2% to 3%, of the clones contain inserts that appear to be highly homologous to sequences from salmonid fishes, as determined by homology comparisons using BLAST and by Southern hybridization analysis to genomic DNA from salmon (Sigma Chemical Co., St. Louis, MO) under stringent washing conditions. The remainder of the clones appear to contain S. japonicum sequences."

ORIGIN
Query Match      4.8%; Score 63.4; DB 13; Length 531;
Best Local Similarity 49.1%; Pred. No. 2.2e-08;
Matches 196; Conservative 0; Mismatches 201; Indels 2; Gaps 1;

QY 696 AAGTGCCTCAGTGGGCCACTGGCTTATCATCGACATTTGGTGAATCTTCTATTATTA 755
DB 7 AATGTCCTATTAAACCGGTGCGGTAAGCCTCGCATCAATGCTGACGGAGATGGGCTCCTG 66

QY 756 CAAACGACACAGCCATCCAGAGGTTCCTGGCAACGCTAAAGGGCGGAGNATGG 815
DB 67 CAACAGACACAGGAATCCGGAGCCGGGTGGGAAGGATCACCGATTCACACAGTGG 126

QY 816 GTATCCTTTATTTGCTATTACATAACGCGCAATTTAATTTAATGGCAAAACACCGTATATC 875
DB 127 AACACCTTGTGAAGTTTGCTAAACGCGCAATTTTATTTGCTACAAACGCGCAGAGGTT 186

QY 876 GCCGCTCATAAGGAGCGCATTTATCAGCAGATAGATACGGCTTTAACCTTCACTG 935
DB 187 GCCCGCAGCGCGCCACCCGGTATTTAGATTGATCAAGACAGCGTGTACGC - CCGGTT 244

QY 936 GATGCTCAGGGGCAAAAGCTACCCATTTTCAGCCAAAACCGGGTCTTGTCTCGGTGGG 995
DB 245 TATCATTTGATTTCATGTTGGAGAGACATTTCCCTCTTGATGACGAAAAGGCGGC 304

QY 996 CATGATACCAATATTGCCMAATTTGGGGTATGCTGGAGCGCATGCGACGCTACCGAG 1055
DB 305 CACGAGCCTGAGCTGATAAACTCGCGCGCGCACTGGAGCTCAACTGACGCTGTCCAGG 364

QY 1056 CAACCTGATAATCTCCACAGGTGGGGATTTGGTTTTT 1094
DB 365 CAGCCCGATACACGCCCGCGCAGGACGGAACTGGATGTT 403

RESULT 7
AW036132
LOCUS

EST274508 tomato seed, TAMU Lycopersicon esculentum cDNA clone
cLEEE223 similar to periplasmic phosphoanhydride phosphohydrolase
precursor, putative, mRNA sequence.
AW036132
EST.
AW036132.1 GI:5894811
Lycopersicon esculentum (tomato)
Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
1 (bases 1 to 383)
Alcala, J., Vrebalov, J., White, R., Matern, A.L., Vision, T.,
Holt, I.E., Liang, F., Upton, J., Ronning, C.M., Craven, M.B.,
Fujii, C.Y., Bowman, C.L., Nierman, W., Fraser, C.M., Venter, J.C.,
Martin, G.B., Tanksey, S.D. and Giovannoni, J.
Generation of ESTs from tomato seed tissue
Unpublished (1999)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
3 prime sequence.
Location/Qualifiers
1. 383
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEEE223"
/tissue_type="seeds"
/dev_stage="quiescent seed"
/lab_host="XLI-Blue MRF"
/clone_lib="tomato seed, TAMU"
/notes="Vector: pBlueScript SK(-); Site 1: EcoRI; Site 2: XhoI; cLEE - Tomato Seed EST Library. Directionally cloned cDNAs inserted into pBlueScript SK(-) at 5' end with EcoRI and 3' end with XhoI site."

ORIGIN
Query Match      4.7%; Score 62.6; DB 9; Length 383;
Best Local Similarity 49.8%; Pred. No. 3.2e-08;
Matches 159; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

QY 365 AACCGACAGGCAATTTCTGTGTGGTTCGGTTCGGGGTGTGGTTTACCGTGCATATCA 424
DB 4 AACAGGGAAGCCCTTCGCGCGCGCTGCGACCTGTGCAATAACCGTACATACCA 63

QY 425 GCGCGATTTGAAAAGACCGATCCCTCTGTTCACCTCGGTAGAGACTGGCGTGTGAAGTT 484
DB 64 GGCAGATACGTCCTCCAGTCCCGATCCGTTATTTAATCTCTAAAACCTGGCGTTGCCACT 123

QY 485 AGACAACGCAACAAAGATAAGCGATTGAAGAACGATTGGGCGGCCCTTAGATACGGT 544
DB 124 GGATAACGGCAACGTGACTGACGCGATCCTCAGCAGGCGAGGAGGTCAATTTGCTGACTT 183

QY 545 GAGCAGCGCTACGCTAAACCTTTTCCGAGATGGGGACGTGCTCAATTTTSCGGCTTC 604
DB 184 TACGGGCATCGCAACCGCGTTTCGGAACCTGGAACCGGTCTTAATTTTCCGCAATC 243

QY 605 TCCTTATTGCAATCTTTGCCAACAGCAAGAAAAACGTTGATTTTGGCCACTTTCCGGC 664
DB 244 AAACCTTGTGCTTAAACGTGAGAAACAGGACGAAAGCTGTTTCATTAAACGAGCATTACC 303

QY 665 CANTGAGTTAAGCTTA 681
DB 304 ATCGGAACCTCAGGTGA 320

RESULT 8
AW036134/c
LOCUS

EST274508 tomato seed, TAMU Lycopersicon esculentum cDNA clone
cLEEE223 similar to periplasmic phosphoanhydride phosphohydrolase
precursor, putative, mRNA sequence.
AW036132
EST.
AW036132.1 GI:5894811
Lycopersicon esculentum (tomato)
Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
1 (bases 1 to 383)
Alcala, J., Vrebalov, J., White, R., Matern, A.L., Vision, T.,
Holt, I.E., Liang, F., Upton, J., Ronning, C.M., Craven, M.B.,
Fujii, C.Y., Bowman, C.L., Nierman, W., Fraser, C.M., Venter, J.C.,
Martin, G.B., Tanksey, S.D. and Giovannoni, J.
Generation of ESTs from tomato seed tissue
Unpublished (1999)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
3 prime sequence.
Location/Qualifiers
1. 383
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEEE223"
/tissue_type="seeds"
/dev_stage="quiescent seed"
/lab_host="XLI-Blue MRF"
/clone_lib="tomato seed, TAMU"
/notes="Vector: pBlueScript SK(-); Site 1: EcoRI; Site 2: XhoI; cLEE - Tomato Seed EST Library. Directionally cloned cDNAs inserted into pBlueScript SK(-) at 5' end with EcoRI and 3' end with XhoI site."

ORIGIN
Query Match      4.8%; Score 63.4; DB 13; Length 531;
Best Local Similarity 49.1%; Pred. No. 2.2e-08;
Matches 196; Conservative 0; Mismatches 201; Indels 2; Gaps 1;

QY 696 AAGTGCCTCAGTGGGCCACTGGCTTATCATCGACATTTGGTGAATCTTCTATTATTA 755
DB 7 AATGTCCTATTAAACCGGTGCGGTAAGCCTCGCATCAATGCTGACGGAGATGGGCTCCTG 66

QY 756 CAAACGACACAGCCATCCAGAGGTTCCTGGCAACGCTAAAGGGCGGAGNATGG 815
DB 67 CAACAGACACAGGAATCCGGAGCCGGGTGGGAAGGATCACCGATTCACACAGTGG 126

QY 816 GTATCCTTTATTTGCTATTACATAACGCGCAATTTAATTTAATGGCAAAACACCGTATATC 875
DB 127 AACACCTTGTGAAGTTTGCTAAACGCGCAATTTTATTTGCTACAAACGCGCAGAGGTT 186

QY 876 GCCGCTCATAAGGAGCGCATTTATCAGCAGATAGATACGGCTTTAACCTTCACTG 935
DB 187 GCCCGCAGCGCGCCACCCGGTATTTAGATTGATCAAGACAGCGTGTACGC - CCGGTT 244

QY 936 GATGCTCAGGGGCAAAAGCTACCCATTTTCAGCCAAAACCGGGTCTTGTCTCGGTGGG 995
DB 245 TATCATTTGATTTCATGTTGGAGAGACATTTCCCTCTTGATGACGAAAAGGCGGC 304

QY 996 CATGATACCAATATTGCCMAATTTGGGGTATGCTGGAGCGCATGCGACGCTACCGAG 1055
DB 305 CACGAGCCTGAGCTGATAAACTCGCGCGCGCACTGGAGCTCAACTGACGCTGTCCAGG 364

QY 1056 CAACCTGATAATCTCCACAGGTGGGGATTTGGTTTTT 1094
DB 365 CAGCCCGATACACGCCCGCGCAGGACGGAACTGGATGTT 403

RESULT 7
AW036132
LOCUS

```

DEFINITION EST274510 tomato seed, TAMU Lycopersicon esculentum cDNA clone  
 cLEE1823 similar to periplasmic phosphoanhydride phosphohydrolase  
 precursor, putative, mRNA sequence.  
 AWO36134  
 AWO36134.1 GI:5894813  
 EST.  
 Lycopersicon esculentum (tomato)  
 Lycopersicon esculentum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.  
 1 (bases 1 to 383)  
 Alcalá, J., Vrebalov, J., White, R., Matern, A.L., Vision, T.,  
 Holt, I.E., Liang, P., Upson, J., Ronning, C.M., Craven, M.B.,  
 Fujii, C.Y., Bowman, C.L., Niernan, W., Fraser, C.M., Venter, J.C.,  
 Martin, G.B., Tanksley, S.D. and Giovannoni, J.  
 Generation of ESTs from tomato seed tissue  
 Unpublished (1999)  
 Contact: CUGI  
 Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Email: <http://www.genome.clemson.edu/orders/index.html>  
 5 prime sequence.  
 Location/Qualifiers  
 1..383  
 /organism="Lycopersicon esculentum"  
 /mol\_type="mRNA"  
 /cultivar="TM496"  
 /db\_xref="taxon:4081"  
 /clone="cLEE1823"  
 /tissue\_type="seeds"  
 /dev\_stage="quiescent seed"  
 /lab\_host="XU1-Blue MRP"  
 /clone\_lib="tomato seed, TAMU"  
 /note="vector: pBlueScript SK(-); Site 1: EcoRI; Site 2:  
 XhoI; cLEE - Tomato seed EST library. Directionally cloned  
 cDNAs inserted into pBlueScript SK(-) at 5' end with  
 EcoRI and 3' end with XhoI site."  
 ORIGIN  
 Query Match 4.7%; Score 62.6; DB 9; Length 383;  
 Best Local Similarity 49.8%; Pred. No. 3.2e-08;  
 Matches 158; Conservative 0; Mismatches 159; Indels 0; Gaps 0;  
 QY 365 AACGGACAGGCATTCTTGATGTGTGGTCCGGGTGTGGTTGACCGTGCAATCA 424  
 DB 380 AACAGGCGAAGCCTTCGCGCGGGCTGGCACCTGACTGTGCATACCGTACATCCCA 321  
 QY 425 GGCCGATTGAAAGAACCGATCCCTGTTCATCCGTAGAGACTGGCGTGTGAAGTT 484  
 DB 320 GGCGATACGTCCAGTCCCGATCCGTTATTAAATCCCTTAAAACTGGCGTTGCCAACT 261  
 QY 485 AGACAACGCAACAACAGATAAAGCGATTGAAGACGATTGGCGGGCGCGTTAGATACGGT 544  
 DB 260 GGATAACGCGAACTGTGACTACCGGATCCTCAGCAGGGCAGGAGGTCAATTGCTGACTT 201  
 QY 545 GAGCAGCGCTACGCTAAACCTTTTGCCAGATGGGGAGCGTGTGAATTTTGGCGTTC 604  
 DB 200 TACGGGCAATCGGCAACCGCGTTTCGCGAACTGGAACGGGTGCTTAATTTTCGCAATC 141  
 QY 605 TCCTTATTGCAAATCTTTTGCACAGCAAGGAAAAACGTGTGATTTTCCCACTTTGGCGG 664  
 DB 140 AAACCTTGCTTAAACGTGAGAAACAGGACGAAGCTGTTCATTACGCGAGGATPACC 81  
 QY 665 CAATGAAGTTAAAGTTA 681  
 DB 80 ATCGGAACCTCAAGGTGA 64  
 RESULT 9  
 BX425603  
 LOCUS

RESULT 10  
BJ619443  
LOCUS  
DEFINITION  
ACCSSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BJ619443 707 bp mRNA linear EST 01-OCT-2003  
Xenopus laevis NIBB Mochii normalized Xenopus early gastrula library  
Xenopus laevis cDNA clone XL189d24 5', mRNA sequence.  
BJ619443  
BJ619443.1 GI:37258203  
Xenopus laevis (African clawed frog)  
Xenopus laevis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
Xenopodinae; Xenopus.  
1 (bases 1 to 707)  
Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and  
Kohara,Y.  
Expressed genes in X. laevis embryo  
Unpublished (2001)  
Contact: Tadasu Shin-i  
Center for Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tehinigenes.nig.ac.jp  
The information of this clone is available through the following  
URL.  
http://xenopus.nibb.ac.jp.

FEATURES  
source  
Location/Qualifiers  
1..707  
/organism="Xenopus laevis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8355"  
/clone="XL189d24"  
/tissue\_type="whole embryo"  
/dev\_stage="stage 10.5"  
/clone\_lib="NIBB Mochii normalized Xenopus early gastrula  
library"

ORIGIN  
Query Match 3.7%; Score 49.2; DB 12; Length 707;  
Best Local Similarity 52.4%; Pred. No. 0.00091;  
Matches 133; Conservative 0; Mismatches 118; Indels 3; Gaps 1;  
QY 771 ATGCCAGAGTTCCTGGCAACGCTAAAGGGGGGAGATTTGGGTATTCCTATTTGCTCA 830  
DB 2 ATCCGGAGCGCGGGTGGGAAGATCACCGATTACACGATTCACCGATTCGTAAGT 61  
QY 831 TTACATAACCGCAATTTAATTTAATGGCAAAACACCGTATATCCCGTCAATTAAGG 890  
DB 62 TTGCATAACCGCAATTTAATTTAATGGCTACACGCAACGAGGTTCCCGCGCGCGCC 121  
QY 891 ACCGCAATTATACAGCAGATAGATACGCTTTAACCCTTCACTGATGCTCAGGGGCAA 950  
DB 122 ACCCGTTATTAGTTGATCAAGACAGCGTTGACGCCCATC---CACCAGCAAAACAG 178  
QY 951 AAGCTACCCATTTTCAGCCCAAAACCGGGTCTTCTTCCTGGTGGGATGATACCAATATT 1010  
DB 179 GCGTATGGTGTGACATTACCCACTTCTAGTGTCTTTATTCGCGGACACGATTAATCTG 238  
QY 1011 GCCAATATTCGGG 1024  
DB 239 GCAAAATCTGGCGG 252

RESULT 11  
BM409016/c  
LOCUS  
DEFINITION  
ACCSSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BM409016 208 bp mRNA linear EST 22-JAN-2002  
EST583343 tomato breaker fruit Lycopersicon esculentum cDNA clone  
CLEG46H24 5' end, mRNA sequence.  
BM409016  
BM409016.1 GI:18260646  
Lycopersicon esculentum (tomato)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.  
1 (bases 1 to 208)  
Alcala,J., Vrebalov,J., White,R., Vision,T., Karamycheva,S.A.,  
Tsai,J., Bougri,O., Kirkness,E., Utterback,T., Van Aken,S.,

SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
source  
Location/Qualifiers  
1..208  
/organism="Lycopersicon esculentum"  
/mol\_type="mRNA"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="CLEG46H24"  
/tissue\_type="pericarp"  
/dev\_stage="breaker"  
/lab\_host="SOLR"  
/clone\_lib="tomato breaker fruit"  
/note="Vector: pBluescriptSKMCluadapt; Site 1: EcoRI;  
Site 2: XhoI; supplier: Boyce Thompson Institute;  
sequencing: The Institute for Genomic Research. Fruit  
were harvested at the breaker stage (first sign of  
lycopene accumulation on the blossom end of fruit). Fruit  
were cut in half and the seeds and locules were discarded  
prior to freezing the pericarp."

ORIGIN  
Query Match 3.7%; Score 48.8; DB 12; Length 208;  
Best Local Similarity 57.1%; Pred. No. 0.00062;  
Matches 89; Conservative 0; Mismatches 67; Indels 0; Gaps 0;  
QY 305 GGGATGTCGCGACAGGGGGGGTATATGACAGGAGATATCGATCAAGTACCCGTT 364  
DB 159 GGGCTGCCCGCAGTCTGGTCAGGTCGCGATTTATTTGCTGACAGGCGTACCCGTAA 100  
QY 365 AACCGGACAGGCATTTCTTGATGTGTGGTCCGGGGTGTGGTGTGACCGTGCAATCA 424  
DB 99 AACAGCGAAGCCTTCGCGCGCGGCTGGCACCTGACTGTGCATATACCGTACATACCA 40  
QY 425 GCGCGATTGAAAAGACCGATCCCTGTTCATCC 460  
DB 39 GCGCAGATACGTCCTCAGTCCCGATCCGTTATTTAATCC 4

RESULT 12  
BM412806/c  
LOCUS  
DEFINITION  
ACCSSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BM412806 208 bp mRNA linear EST 22-JAN-2002  
EST587133 tomato breaker fruit Lycopersicon esculentum cDNA clone  
CLEG61013 5' end, mRNA sequence.  
BM412806  
BM412806.1 GI:18264436  
Lycopersicon esculentum (tomato)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.  
1 (bases 1 to 208)  
Alcala,J., Vrebalov,J., White,R., Vision,T., Karamycheva,S.A.,  
Tsai,J., Bougri,O., Kirkness,E., Utterback,T., Van Aken,S.,

Roanling, C.M., Fraser, C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, J.  
Generation of ESTs from tomato fruit tissue, breaker stage (2002)  
Unpublished (2002)  
Contact: CUGI

Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>  
This clone is available through the Clemson University Genomics Institute

## Seq primer: T3

## Location/Qualifiers

1. 208  
/organism="Lycopersicon esculentum"  
/mol\_type="mRNA"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="CLSG61013"  
/tissue\_type="pericarp"  
/dev\_stage="breaker"  
/lab\_host="SOLR"  
/clone\_lib="tomato breaker fruit"  
/note="Vector: pBluescriptSKmCUadapt; Site 1: EcoRI; Site 2: XhoI; supplier: Boyce Thompson Institute; sequencing: The Institute for Genomic Research. Fruit were harvested at the breaker stage (first sign of lycopene accumulation on the blossom end of fruit). Fruit were cut in half and the seeds and locules were discarded prior to freezing the pericarp."

## ORIGIN

Query Match 3.7%; Score 48.8; DB 12; Length 208;  
Best Local Similarity 57.1%; Pred. No. 0.00062;  
Matches 89; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 305 GGGATGTCGGCAGAGGGGGGTATATGACAGGAGATATCGATCAAGTACCGCTT 364  
Db 159 GGGCTGCCCGAGCTGCTGAGTGGCGGATATTGCTGATGTCAGGAGGTACCGTAA 100  
QY 365 AACCGGACGAGCATTTCTTGATGTGTGGCTCCGGGTGGTTGACCGTGCATATCA 424  
Db 99 AACGCGAAGCCTTTCGGCCGGCGCTGGACCTGTGCAATACCGTACATACCA 40  
QY 425 GCGCATTTGAAAAGACGATCCCTTTCATCC 460  
Db 39 GCGAGATACGTCACATCCGATCCGTTATTATCC 4

RESULT 13  
AA545747/c  
LOCUS  
DEFINITION  
HBM5F1B4-REV Human Bone Marrow Stromal Fibroblast Homo sapiens cDNA  
clone HBM5F1B4 5', mRNA sequence.  
AA545747  
ACCESSION  
AA545747.1 GI:2307026  
VERSION  
KEYWORDS  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 354)  
Jia, L., Robey, P., Young, M., Schuler, G., Powell, J., Yang, L.,  
Lennon, G., Hillier, L., Allen, M., Bowles, L., Geisel, S., Kucaba, T.,  
Marra, M., Martin, J., Steptoe, M., Tan, F., Theising, B., Bowers, Y.,  
Wylie, T., Waterson, R., Wilson, R. and Francomano, C.  
WASHU-MGB/NHGR1 EST Project  
Unpublished (1997)  
Contact: Libin Jia  
Medical Genetics Branch  
National Human Genome Research Institute  
10/10C101, 9000 Rockville Pike, Bethesda, MD 20892-1267, USA  
Tel: 301-402-4877

FEATURES  
source

FEATURES  
source  
1. 354  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="HBM5F1B4"  
/sex="Male and Female"  
/tissue\_type="bone marrow"  
/cell\_type="stromal fibroblast"  
/dev\_stage="mixed"  
/lab\_host="XL1-Blue"  
/clone\_lib="Human Bone Marrow Stromal Fibroblast"  
/note="Vector: pBluescript; Site 1: EcoRI; Site 2: XhoI"

## ORIGIN

Query Match 3.5%; Score 46.2; DB 9; Length 354;  
Best Local Similarity 53.8%; Pred. No. 0.0057;  
Matches 93; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 696 AAGTGCACCTCAGTGGGCGCTATCATCATGCGTGGTGAATCTTCTATTATTA 755  
Db 174 ATGTTTCATTACCGGTGGGTAGCTCGCATCAATGCTGACGAGATATTTCCTG 115  
QY 756 CAAAGCGCAAGCCATGCCAGAGTTGCTGCGCAACGGCTAAAGGGGGGAGAAATGG 815  
Db 114 CAACAGCACAGGAATCGGAGCGGGTGGGAGGATCACCGATTACACAGTGG 55  
QY 816 GTATCCTTATTCTCATTAACCGCAATTTAATTTAATGCGAAAAACACC 868  
Db 54 AACACCTTGCTAAGTTGCATACGCGCAATTTATTGCTACAACGACGCGC 2

RESULT 14  
BX381961/c

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

BX381961 1201 bp mRNA linear EST 08-MAY-2003  
clone CSOD1072YF05 3-PRIME, mRNA sequence.  
BX381961  
BX381961.1 GI:30453007  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1201)  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: [seqref@genoscope.cns.fr](mailto:seqref@genoscope.cns.fr), Web: [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Library was constructed by Life Technologies, a division of  
Invitrogen. Contact: Feng Liang Email: [fliang@lifetech.com](mailto:fliang@lifetech.com) URL:  
<http://fulllength.invitrogen.com/> Invitrogen Corporation 1600  
Paradise Avenue Genoscope sequence ID: CSOD1072CC03NP1.

FEATURES  
source

1. 1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSOD1072YF05"  
/tissue\_type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with NotI and cloned into the NotI and EcoRV  
sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN



Query Match 3.3%; Score 44; DB 13; Length 1201;  
Best Local Similarity 4.5%; Pred.No.0.057;  
Matches 31; Conservative 245; Mismatches 418; Indels 1; Gaps 1;  
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org>. The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Oosagawa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES  
source 1..1101  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/clone="BAC17F22"  
/clone\_lib="RPCI-98"  
/notes="end : T7"

ORIGIN

Query Match 3.3%; Score 43.2; DB 29; Length 1101;  
Best Local Similarity 10.1%; Pred.No.0.098;  
Matches 36; Conservative 161; Mismatches 160; Indels 0; Gaps 0;  
QY 82 GCGCGGAGCATCGGGCTATCTTTAGAACGGTGGTGGTATTTAGTGGCCATGGTGT 141  
DB 1090 KCHNKK 1031  
QY 142 CGTTCGCGGACCAACACGACGCTTATGATGATGTTAGCCAGATAGTGCCACAT 201  
DB 1030 MKKK 971  
QY 202 GCGCGGTAAGGCGGGTATTTAAACGCGCTGGTGGGAGTTGGTCCATTCATGGGG 261  
DB 970 KMKKCCCTKKCKGCKGCKTKCTKCHNMTKSTCBMTKTKKVKGMATKMMGCKGA 911  
QY 262 GGTTTTATGTTGATTAATTTTCGACGCTTGGTTTGTAGCGGGGATGTCGCGCAGAG 321  
DB 910 CKDGKKMKDBKMDKK 851  
QY 322 GGGGGGTATGACAGGCGAGATATCGATCAACGTACCGCTTAAACGCGACGCGATTTC 381  
DB 850 KKKKTKTKMMTKKMMTKKMMTKKMMTKKMMTKKMMTKKMMTKKMMTKKMMTKKMM 791  
QY 382 TTGATGCTGGTGGTCCGGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 438  
DB 790 TMTKVVMTTK 734

Search completed: May 3, 2004, 03:54:06  
Job time : 2502.04 secs

QY 53 GCGGATTCGGCTATTACTGCGCGGTAGCGCGGAGCCATCGGGTATCTTTAGAAC 112  
DB 1087 GKKGCGSKGKVKMGKMMKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 1028  
QY 113 GTGTGTTATTTAGTCCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 171  
DB 1027 KKKTKKKKMTGKK 968  
QY 172 ATGATGTTACGCGAGATGAAGTGGCCACATGCGCGGTAAAGCGGGGTATTTAAACG 231  
DB 967 VKKBMGGGKMKKMKKMGKMMKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 908  
QY 232 GTGTGCGGAGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 291  
DB 907 KKKGKMMKK 848  
QY 292 GTTTCGCGGCGGATGTCGCGGAGGCGGGGTATATGACAGGCGAGATATCGATC 351  
DB 847 MYGKMKGKMKKVKMGKMMKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 788  
QY 352 AACGTACCGGCTTAACCGGACAGGCAATTTCTTGTGCTGCTGCTGCTGCTGCTG 411  
DB 787 MCKMMNN 728  
QY 412 CGTGCATATCAGCGCGATTTGAAAGACCGATCCCTGTTCCATCGGTAGAGACTG 471  
DB 727 NNANNNTNN 668  
QY 472 GCGTGTGTAAGTTAGACACGACAAAGAGATTAAGAGATTAAGAGATTAAGAGAT 531  
DB 667 NNN 608  
QY 532 CGTTAGATACGCTGAGCGAGCGCTACGCTAAACCTTTTCCGAGATGGGGAGCTG 591  
DB 607 NNNNNNTNN 548  
QY 592 ATTTCGCGCTCTCTCTTATTCGAAATCTTTGCAACGAGGAGGAGGAGGAGGAG 651  
DB 547 KKKKKMMNNNTKCKGNN 488  
QY 652 CCCATTTTCGCGCAATGAGTTAAGTTAAGTTAAGTTAAGTTAAGTTAAGTTAAG 711  
DB 487 MNN 428  
QY 712 GCGCATGCGCTTATCATCGACATTCGCTGAAATC 746  
DB 427 KMMNN 393

RESULT 15  
CNS00KK2/c  
LOCUS  
DEFINITION CNS00KK2 1101 bp DNA linear GSS 03-JUN-1999  
Drosophila melanogaster genome survey sequence T7 end of BAC:  
BAC17F22 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
ACCESSION AL077673  
VERSION AL077673.1 GI:4957249  
KEYWORDS GSS.  
SOURCE Drosophila melanogaster (fruit fly)  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 1101)  
Genoscope.  
REFERENCE Direct Submission  
AUTHORS Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
TITLE BP 191 91006 ERY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
JOURNAL - Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr))



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 12:30:35 ; Search time 53.7805 Seconds  
(without alignments)  
2206.561 Million cell updates/sec

Title: US-10-021-723B-4  
Perfect score: 2200  
Sequence: 1 MSVLENRVRLSGLVLMISGL.....NPAGIIISVAVACENNGDDK 420

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04.\*

- 1: Geneseq1980s.\*
- 2: Geneseq1990s.\*
- 3: Geneseq2000s.\*
- 4: Geneseq2001s.\*
- 5: Geneseq2002s.\*
- 6: Geneseq2003as.\*
- 7: Geneseq2003bs.\*
- 8: Geneseq2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2200	100.0	420	5 ABP51930	Abp51930 Yersinia
2	2193	99.7	441	5 ABP51934	Abp51934 Yersinia
3	2183	99.2	441	6 ABU49868	Abu49868 Protein e
4	2175	98.9	441	5 ABP51929	Abp51929 Yersinia
5	957.5	43.5	431	6 AAE37853	Aae37853 PNOV4054
6	957.5	43.5	437	6 AAE37854	Aae37854 PNOV4058
7	948.5	43.1	436	7 ADC87743	Adc87743 Modified
8	943	42.9	410	5 ABP51937	Abp51937 Phytase p
9	943	42.9	412	6 AAE37851	Aae37851 Maize-opt
10	939.5	42.7	432	7 ABR82310	AbR82310 Nov9X phy
11	939.5	42.7	432	3 AAB36257	Aab36257 Lama2/APP
12	939.5	42.7	432	3 AAB36261	Aab36261 R15/APP
13	939.5	42.7	432	3 AAB36262	Aab36262 SV40/APP
14	939.5	42.7	432	3 AAB36259	Aab36259 R15/APP
15	939.5	42.7	432	3 AAB36258	Aab36258 R15/APP
16	939.5	42.7	432	3 AAB36263	Aab36263 Lama2/APP
17	939.5	42.7	432	3 AAB36260	Aab36260 R15/APP
18	939.5	42.7	432	3 AAU77775	Aau77775 Phytase a
19	939.5	42.7	432	5 AAE15807	Aae15807 Escherich
20	939.5	42.7	432	5 ABP53625	Abp53625 Acid phos
21	939.5	42.7	432	6 ADAL19450	Adal19450 E. coli K
22	939.5	42.7	432	6 ABU15427	Abu15427 Protein e
23	939.5	42.7	432	7 ADC87745	Adc87745 Escherich
24	938.5	42.7	432	4 AAE02631	Aae02631 E. coli a
25	938.5	42.7	432	4 AAE02634	Aae02634 E. coli a

## ALIGNMENTS

### RESULT 1

ABP51930

ID ABP51930 standard; protein; 420 AA.

XX AC ABP51930;

DT 08-OCT-2002 (first entry)

XX Yersinia pestis phytase protein sequence SEQ ID NO:4.

DE Yersinia pestis; phytase; enzyme; E.C.3.1.3.8; phytate; foodstuff; feed.

XX Yersinia pestis.

OS Yersinia pestis.

XX WO200248332-A2.

XX 20-JUN-2002.

XX 12-DEC-2001; 2001WO-US048774.

XX 12-DEC-2000; 2000US-0255090P.

XX (DIVE-) DIVERSA CORP.

XX Short J, Mathur EJ, Richardson T, Robertson D, Barton N;

XX WPI; 2002-583504/62.

XX N-PSDB; ABQ73843.

XX Novel recombinant phytase protein and polynucleotide for improving nutritional value of phytate-containing foodstuff, in animal feed and feed supplements and to degrade excess phytase from environment or sample.

XX Claim 49; Fig 5D; 208pp; English.

The present invention describes an isolated phytase protein (I). (I) can be used for improving the nutritional value of a phytate-containing foodstuff, by contacting the phytate-containing foodstuff with (I), where the enzyme catalyses the liberation of inorganic phosphate from the phytate-containing foodstuff, and so improving the nutritive value of the contacted foodstuff. The liberation of inorganic phosphate occurs prior to or after the ingestion of phytate-containing foodstuff by a recipient organism. Nucleotide sequences (II) encoding (I) can be used for producing an animal feed, by transforming a plant, plant portion or plant cell with a nucleic acid expression vector, comprising (II), culturing the plant, plant portion or plant cell under conditions in which the phytase protein is expressed and converting the plant, plant portion or

Aab37892 Escherich  
Aae22836 Escherich  
Aae15806 Escherich  
Adal19446 E. coli B  
Aae02635 E. coli a  
Aay24753 E. coli a  
Aae02632 E. coli a  
Abr42161 Escherich  
Abp1933 Yersinia  
Aae15808 Escherich  
Adal19451 E. coli K  
Aay01513 An Escher  
Aab51936 Consensus  
Abp1932 Rhizobium  
Abp1931 Yersinia  
Abg25698 Novel hum  
Aau23269 Novel hum  
Abp51935 Appa phyt  
Abg24256 Novel hum

26 932.5 42.4 440 4 AAB37892  
27 932.5 42.4 440 5 AAE22836  
28 932.5 42.4 440 5 AAE15806  
29 932.5 42.4 440 6 ADAL19446  
30 931.5 42.3 432 4 AAE02635  
31 928.5 42.2 432 3 AAY94753  
32 928.5 42.2 432 4 AAE02632  
33 928.5 42.2 432 6 ABR42161  
34 928.5 42.2 476 5 ABP51933  
35 921.5 41.9 430 5 AAE15808  
36 920.5 41.8 432 6 ADAL19451  
37 860 39.1 423 2 AAY01513  
38 567 25.8 318 5 ABP51936  
39 504 22.9 409 5 ABP51932  
40 482 21.9 421 5 ABP51931  
41 469 21.3 261 4 ABG25698  
42 291 13.2 123 4 AAU23269  
43 282 12.8 393 5 ABP51935  
44 265 12.0 144 4 ABG24256

CC plant cell into a composition suitable for animal feed. The animal is  
CC preferably a monogastric animal or a ruminant. (I) and (II) are useful in  
CC animal feed and feed supplements as well as in treatments to degrade or  
CC remove excess phytate from the environment or a sample. (I) reduces  
CC phytate levels in animal manure and so reduces the phosphate pollution of  
CC the environment. The present sequence represents a phytase from the  
CC present invention  
XX  
SQ Sequence 420 AA;

Query Match 100.0%; Score 2200; DB 5; Length 420;  
Best Local Similarity 100.0%; Pred. No. 1.8e-210;  
Matches 420; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MSVLENRVRLSGLVLMISGLAAITAPVAAPSPGYTLRRVILSRHGVRSPTKOTQLMNDV 60  
Db 1 MSVLENRVRLSGLVLMISGLAAITAPVAAPSPGYTLRRVILSRHGVRSPTKOTQLMNDV 60  
QY 61 TPDKWPQWPVKAGVLTFRGAEVLTLMGFFYDIFRSGLLAACCPAEGGVYAQADIDQRT 120  
Db 61 TPDKWPQWPVKAGVLTFRGAEVLTLMGFFYDIFRSGLLAACCPAEGGVYAQADIDQRT 120  
QY 121 RLGTQAFLDGVAPEGGLTVHNAQDLKKTDLFHPVETGVCKLNAQTDKAIERLGGPLD 180  
Db 121 RLGTQAFLDGVAPEGGLTVHNAQDLKKTDLFHPVETGVCKLNAQTDKAIERLGGPLD 180  
QY 181 TVSORYAKPQAQMGDVILNFAASPYCKSLQOQGTCTDFAHFAANVNNKEGTVTLGGPL 240  
Db 181 TVSORYAKPQAQMGDVILNFAASPYCKSLQOQGTCTDFAHFAANVNNKEGTVTLGGPL 240  
QY 241 ALSSTLGEIFLLQNAQAMPVAVQRLKGAENWVSLSLHNAQFNLMAKTPYIARHKGTEL 300  
Db 241 ALSSTLGEIFLLQNAQAMPVAVQRLKGAENWVSLSLHNAQFNLMAKTPYIARHKGTEL 300  
QY 301 LQOITDALTQLDAQOGKLPISQNRVFLFGGHDNTNANTAGMLGADWQLPEQPDNTPPG 360  
Db 301 LQOITDALTQLDAQOGKLPISQNRVFLFGGHDNTNANTAGMLGADWQLPEQPDNTPPG 360  
QY 361 GGLVFELWQPNQPNHQRVAVVMFYQTMQDLRNEAKLDLKNPAGIISVAVACENNGDDK 420  
Db 361 GGLVFELWQPNQPNHQRVAVVMFYQTMQDLRNEAKLDLKNPAGIISVAVACENNGDDK 420

RESULT 2  
ABP51934  
ID ABP51934 standard; protein; 441 AA.  
XX AC ABP51934;  
XX DT 08-OCT-2002 (first entry)  
XX DE Yersinia pestis phytase protein sequence.  
XX KW Yersinia pestis; phytase; enzyme; E.C.3.1.3.8; phytate; foodstuff; feed.  
XX OS Yersinia pestis.  
XX FN W0200248332-A2.  
XX PD 20-JUN-2002.  
XX PF 12-DEC-2001; 2001WO-US048774.  
XX PR 12-DEC-2000; 2000US-0255090P.  
XX PA (DIVE-) DIVERSA CORP.  
XX PI Short J, Mathur EJ, Richardson T, Robertson D, Barton N;  
XX WPI: 2002-583504/62.  
XX DR N-PSDB; ABQ73847.  
XX PT Novel recombinant phytase protein and polynucleotide for improving

PT nutritional value of phytate-containing foodstuff, in animal feed and  
PT feed supplements and to degrade excess phytase from environment or  
PT sample.  
XX  
XX Claim 57; Fig 5L; 208pp; English.

CC The present invention describes an isolated phytase protein (I). (I) can  
CC be used for improving the nutritional value of a phytate-containing  
CC foodstuff, by contacting the phytate-containing foodstuff with (I), where  
CC the enzyme catalyses the liberation of inorganic phosphate from the  
CC phytate-containing foodstuff, and so improving the nutritive value of the  
CC contacted foodstuff. The liberation of inorganic phosphate occurs prior  
CC to or after the ingestion of phytate-containing foodstuff by a recipient  
CC organism. Nucleotide sequences (II) encoding (I) can be used for  
CC producing an animal feed, by transforming a plant, plant portion or plant  
CC cell with a nucleic acid, by transfection vector, comprising (II), culturing  
CC the plant, plant portion or plant cell under conditions in which the  
CC phytase protein is expressed and converting the plant, plant portion or  
CC plant cell into a composition suitable for animal feed. The animal is  
CC preferably a monogastric animal or a ruminant. (I) and (II) are useful in  
CC animal feed and feed supplements as well as in treatments to degrade or  
CC remove excess phytate from the environment or a sample. (I) reduces  
CC phytate levels in animal manure and so reduces the phosphate pollution of  
CC the environment. The present sequence represents a phytase from the  
CC present invention  
XX

Query Match 99.7%; Score 2193; DB 5; Length 441;  
Best Local Similarity 99.8%; Pred. No. 9.9e-210;  
Matches 419; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MSVLENRVRLSGLVLMISGLAAITAPVAAPSPGYTLRRVILSRHGVRSPTKOTQLMNDV 60  
Db 1 MSVLENRVRLSGLVLMISGLAAITAPVAAPSPGYTLRRVILSRHGVRSPTKOTQLMNDV 60  
QY 61 TPDKWPQWPVKAGVLTFRGAEVLTLMGFFYDIFRSGLLAACCPAEGGVYAQADIDQRT 120  
Db 61 TPDKWPQWPVKAGVLTFRGAEVLTLMGFFYDIFRSGLLAACCPAEGGVYAQADIDQRT 120  
QY 121 RLGTQAFLDGVAPEGGLTVHNAQDLKKTDLFHPVETGVCKLNAQTDKAIERLGGPLD 180  
Db 121 RLGTQAFLDGVAPEGGLTVHNAQDLKKTDLFHPVETGVCKLNAQTDKAIERLGGPLD 180  
QY 181 TVSORYAKPQAQMGDVILNFAASPYCKSLQOQGTCTDFAHFAANVNNKEGTVTLGGPL 240  
Db 181 TVSORYAKPQAQMGDVILNFAASPYCKSLQOQGTCTDFAHFAANVNNKEGTVTLGGPL 240  
QY 241 ALSSTLGEIFLLQNAQAMPVAVQRLKGAENWVSLSLHNAQFNLMAKTPYIARHKGTEL 300  
Db 241 ALSSTLGEIFLLQNAQAMPVAVQRLKGAENWVSLSLHNAQFNLMAKTPYIARHKGTEL 300  
QY 301 LQOITDALTQLDAQOGKLPISQNRVFLFGGHDNTNANTAGMLGADWQLPEQPDNTPPG 360  
Db 301 LQOITDALTQLDAQOGKLPISQNRVFLFGGHDNTNANTAGMLGADWQLPEQPDNTPPG 360  
QY 361 GGLVFELWQPNQPNHQRVAVVMFYQTMQDLRNEAKLDLKNPAGIISVAVACENNGDDK 420  
Db 361 GGLVFELWQPNQPNHQRVAVVMFYQTMQDLRNEAKLDLKNPAGIISVAVACENNGDDK 420

RESULT 3  
ABU49868  
ID ABU49868 standard; protein; 441 AA.  
XX AC ABU49868;  
XX DT 19-JUN-2003 (first entry)  
XX DE Protein encoded by Prokaryotic essential gene #35395.  
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX



CC be used for improving the nutritional value of a phytate-containing  
CC foodstuff, by contacting the phytate-containing foodstuff with (I), where  
CC the enzyme catalyses the liberation of inorganic phosphate from the  
CC phytate-containing foodstuff, and so improving the nutritive value of the  
CC contacted foodstuff. The liberation of inorganic phosphate occurs prior  
CC to or after the ingestion of phytate-containing foodstuff by a recipient  
CC organism. Nucleotide sequences (II) encoding (I) can be used for  
CC producing an animal feed, by transforming a plant, plant portion or plant  
CC cell with a nucleic acid expression vector, comprising (II), culturing  
CC the plant, plant portion or plant cell under conditions in which the  
CC phytase protein is expressed and converting the plant, plant portion or  
CC plant cell into a composition suitable for animal feed. The animal is  
CC preferably a monogastric animal or a ruminant. (I) and (II) are useful in  
CC animal feed and feed supplements as well as in treatments to degrade or  
CC remove excess phytate from the environment or a sample. (I) reduces  
CC phytate levels in animal manure and so reduces the phosphate pollution of  
CC the environment. The present sequence represents a phytase from the  
CC present invention  
XX  
SQ Sequence 441 AA;

Query Match 98.9%; Score 2175; DB 5; Length 441;  
Best Local Similarity 99.0%; Pred. No. 6.2e-208;  
Matches 416; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
Qy 1 MSVLENRVLSGLVLMISGLAATAPVAAPSPGYTLRVVLSRHGVRSPKOTQLMNDV 60  
Db 1 MSVLENRVLSGLVLMISGLAATAPVAAPSPGYTLRVVLSRHGVRSPKOTQLMNDV 60  
Qy 61 TPKWQWPVKAGYLTFRGAEVLTLMGGFYGYFRSLGLLAAGCPAEGGVYQAADIDQRT 120  
Db 61 TPKWQWPVKAGYLTFRGAEVLTLMGGFYGYFRSLGLLAAGCPAEGGVYQAADIDQRT 120  
Qy 121 RLGTQAPLDGAVPGCGLTVHQAADLKKTDPLFHPVETGVCKLDNAOTDKAIEERLGGPLD 180  
Db 121 RLGTQAPLDGAVPGCGLTVHQAADLKKTDPLFHPVETGVCKLDNAOTDKAIEERLGGPLD 180  
Qy 181 TVSORVAKPPAQMAGDVLNPAAPSPYCKSLQQQKTCDFAHFAENNVNKEGTVTLGGPL 240  
Db 181 TVSORVAKPPAQMAGDVLNPAAPSPYCKSLQQQKTCDFAHFAENNVNKEGTVTLGGPL 240  
Qy 241 ALSTLGEIFLLQNAQAMPVAVQRLKGAENWVLSLHNAQFNLMKATPYIARHKGTP 300  
Db 241 ALSTLGEIFLLQNAQAMPVAVQRLKGAENWVLSLHNAQFNLMKATPYIARHKGTP 300  
Qy 301 LQITDTALTQLDAQGQKLPISAQNRVFLGGHDTNIAIAGMLGADWQLEPQDNTPPG 360  
Db 301 LQITDTALTQLDAQGQKLPISAQNRVFLGGHDTNIAIAGMLGADWQLEPQDNTPPG 360  
Qy 361 GGLVFELWQPNDRHQRVAVQMFYQTDQLRNEAKLDLKNPAGIISVAVAGCENGGDK 420  
Db 361 GGLVFELWQPNDRHQRVAVQMFYQTDQLRNEAKLDLKNPAGIISVAVAGCENGGDK 420

RESULT 5  
AAE37853  
ID AAE37853 standard; protein; 431 AA.  
XX  
AC AAE37853;  
XX  
DT 06-NOV-2003 (first entry)  
XX  
DE pNOV4054 phytase fusion protein.  
XX  
KW Thermotolerant phytase; weight gain; animals fed diet; grain processing;  
KW nutritive value; transformed plant; anabolic; chimeric; maize.  
XX  
OS Chimeric - Zea mays.  
OS Chimeric - Unidentified.  
XX  
XX Key Location/Qualifiers  
XX Peptide 1..19  
FT /label= Signal\_peptide

Protein 20. .431  
/note= "Mature pNOV4054 phytase fusion protein"  
XX WO2003057248-A1.  
XX 17-JUN-2003.  
XX 30-DEC-2002; 2002WO-US041787.  
XX 28-DEC-2001; 2001US-0344476P.  
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX Lanahan MB, Betts S;  
XX WPI; 2003-607980/57.  
XX N-PSDB; AADS7148.  
XX Preparing a thermotolerant phytase for preparing animal feed or human  
XX food by expressing in a plant cell an expression cassette comprising a  
XX promoter operably linked to a nucleic acid molecule encoding a  
XX thermotolerant phytase.  
XX Claim 63; Page 90; 157pp; English.  
XX  
XX The invention relates to a method for preparing a thermotolerant phytase.  
XX The method comprises expressing in a plant cell an expression cassette  
XX comprising a promoter operably linked to a nucleic acid molecule encoding  
XX a thermotolerant phytase which retains at least 40% activity after 30  
XX minutes at 60pH and has a specific activity of greater than 200 U/mg  
XX at pH 4.5 and 37pH. The method is useful for preparing a  
XX thermotolerant phytase for preparing animal feed or human food. The  
XX invention is useful for reducing the feed conversion ratio and increasing  
XX weight gain, improving reducing feed conversion ratios or increasing  
XX weight gain of animals fed diets with inorganic phosphate at levels below  
XX 0.45%, minimising dietary requirements of phosphorus in an animal,  
XX enhancing the utilisation of phosphorus present in animal feed, enhancing  
XX organic phosphorus utilisation from organic phosphorus sources in animal  
XX feed, decreasing the phosphate levels in excreta from an animal,  
XX improving the processing of grain, improving the nutritive value of  
XX processed grain product or a method of processing grain, improving the  
XX nutritive value of animal feed and human food, and preparing a  
XX transformed plant which expresses a thermotolerant phytase. The present  
XX sequence is pNOV4054 phytase fusion protein used in the exemplification  
XX of the invention  
XX  
SQ Sequence 431 AA;

Query Match 43.5%; Score 957.5; DB 6; Length 431;  
Best Local Similarity 48.3%; Pred. No. 2.4e-86;  
Matches 199; Conservative 63; Mismatches 137; Indels 13; Gaps 7;  
Qy 7 RVLISGLVLMISGLAATAPVAAPSPGYTLRVVLSRHGVRSPKOTQLMNDVTPDKWP 66  
Db 2 RVLISGLVLMISGLAATAPVAAPSPGYTLRVVLSRHGVRSPKOTQLMNDVTPDKWP 66  
Qy 67 QWPKAGYLTFRGAEVLTLMGGFYGYFRSLGLLAAGCPAEGGVYQAADIDQRTLTGQ 125  
Db 60 TWPVKGLTFRGAEVLTLMGGFYGYFRSLGLLAAGCPAEGGVYQAADIDQRTLTGQ 119  
Qy 126 AFLDGVAPCGTLVHQAADLKKTDPLFHPVETGVCKLDNAOTDKAIEERLGGPLDTSQR 185  
Db 120 AFAAGLAPDCAITVHTQADTSSPDPLFNPLKTVGVCOLDNANVTDAILERAGGSIAFTGH 179  
Qy 186 YAKPFAQMGDVLNPAAPSPYCKSLQQQKTCDFAHFAENNVNKEGTVTLGGPLDTSQR 245  
Db 180 YTAPEFLRVNLPQSNLKRKREKQDECSLTQALPSLKVSA--CVSITGAVSLASM 237  
Qy 246 LGEIIFLLQNAQAMPVAVQRLKGAENWVLSLHNAQFNLMKATPYIARHKGTPLOQID 305  
Db 238 LGEIIFLLQNAQAMPVAVQRLKGAENWVLSLHNAQFNLMKATPYIARHKGTPLOQID 297  
Qy 306 TALT---LQIDAQGQKLPISAQNRVFLGGHDTNIAIAGMLGADWQLEPQDNTPPGG 362

Db 298 TALTTHPPKQKQYGVTLPTS-----VLFAGHDTNLNAGGALNLTWLPQPDNTPGGE 353

QY 363 LVFELWQNPDRHQRYVAVKMFYQTMQDLNRNAEKLDLKNPAGIISVAVAGCE 414

Db 354 LVFERWRRLSDNSQWQVSLVFTLQQRDXTPLSL-NTPPGEVKLTLAGCE 404

RESULT 6

AAE37854

ID AAE37854 standard; protein; 437 AA.

AC AAE37854;

DT 06-NOV-2003 (first entry)

DE pNOV4058 phytase fusion protein.

XX Thermotolerant phytase; weight gain; animals fed diet; grain processing;

KW nutritive value; transformed plant; anabolic; chimeric; maize.

XX Chimeric - Zea mays.

OS Chimeric - Unidentified.

XX

FH Key Location/Qualifiers

FT Peptide 1..19

FT /label= Signal\_peptide

FT Protein 20..437

FT /note= "Mature pNOV4058 phytase fusion protein"

FT Region 432..437

FT /note= "SEXDEL signal"

XX

WO2003057248-A1.

PN

XX

PD 17-JUL-2003.

PF 30-DEC-2002; 2002WO-US041787.

XX

PR 28-DEC-2001; 2001US-0344476P.

XX

PA (SYGN ) SYNGENTA PARTICIPATIONS AG.

XX

PI Lanahan MB, Betts S;

XX

DR WPI; 2003-607980/57.

DR N-PSDB; AAD57149.

XX

PT Preparing a thermotolerant phytase for preparing animal feed or human

PT food by expressing in a plant cell an expression cassette comprising a

PT promoter operably linked to a nucleic acid molecule encoding a

PT thermotolerant phytase.

XX

PS Example 1; Page 91; 157pp; English.

XX

CC The invention relates to a method for preparing a thermotolerant phytase.

CC The method comprises expressing in a plant cell an expression cassette

CC comprising a promoter operably linked to a nucleic acid molecule encoding

CC a thermotolerant phytase which retains at least 40% activity after 30

CC minutes at 60pluoc and has a specific activity of greater than 200 U/mg

CC at pH 4.5 and 37pluoc. The method is useful for preparing a

CC thermotolerant phytase for preparing animal feed or human food. The

CC invention is useful for reducing the feed conversion ratio and increasing

CC weight gain, improving reduced feed conversion ratios or increasing

CC weight gain of animals fed diets with inorganic phosphate at levels below

CC 0.45%, minimising dietary requirements of phosphorus in an animal,

CC enhancing the utilisation of phosphorus present in animal feed, enhancing

CC organic phosphorus utilisation from organic phosphorus sources in animal

CC feed, decreasing the phosphate levels in excreta from an animal,

CC improving the processing of grain, improving the nutritive value of

CC processed grain product or a method of processing grain, improving the

CC nutritive value of animal feed and human food, and preparing a

CC transformed plant which expresses a thermotolerant phytase. The present

CC transference is pNOV4058 phytase fusion protein used in the exemplification

CC of the invention

XX

SQ Sequence 437 AA;

Query Match 43.5%; Score 957.5; DB 6; Length 437;

Best Local Similarity 48.3%; Pred. No. 2.5e-86;

Matches 199; Conservative 63; Mismatches 137; Indels 13; Gaps 7;

QY 7 RVLSGLVLMISGLAAITAPVAAEPGTYLIERVILSRHGVRSPTKQTOLMNDVTPDKWP 66

Db 2 RVLLVALAL-LALASATSAQSEPE-LKLESVIVSRHGVRAPTKATQIMQDVTDPDWP 59

QY 67 QNPVKAGYLTTPRGAEALVTLMGFGFYGDYFRSLGLL-AAGCPAEGGVYAQADIDQRTLTQ 125

Db 60 TWPVKLGELTPRGELIAYLGHVWRQRLVADGLLPGCGPQSQGVAIADVDERTKTGE 119

QY 126 AFLDGVAPCGGLTVRNOADLKTDDPLFHPVETGVCKLDNAQTDKALIEERLGGPLDVSR 185

Db 120 AFAAGLAPDCAITVHTQADTSSPDPLFNELKTVGCQLDNANVTDAILERAGSIADFTGH 179

QY 186 YAKPPAQMGDVNLNFAASPYCKSLQOQKCTDFAHFAANVNVNKEGTKVTLSGLALSST 245

Db 180 YQTAFRELERLVNFQSNLCKREKQDESCSLTQALPSELKVSAD--CVSLTAGVSLASM 237

QY 246 LGEIFLLQNAQAMPEVANQRLKGAENWVLSLHNAQFNLMKATPYIAHKGTPLLQOID 305

Db 238 LTBIFLLQQAQGMPEPGWGRITDSDHWNLTLSLHNAQFDLLQRTPEVARSRATPLLDLX 297

QY 306 TALT---LQDAGQKLPISAQNRVFLGCHDTNIANIAGMLGADWOLPEQDNTPPGG 362

Db 298 TALTTHPPKQKQYGVTLPTS-----VLFAGHDTNLNAGGALNLTWLPQPDNTPGGE 353

QY 363 LVFELWQNPDRHQRYVAVKMFYQTMQDLNRNAEKLDLKNPAGIISVAVAGCE 414

Db 354 LVFERWRRLSDNSQWQVSLVFTLQQRDXTPLSL-NTPPGEVKLTLAGCE 404

RESULT 7

ADC87743

ID ADC87743 standard; protein; 436 AA.

XX

AC ADC87743;

DT 01-JAN-2004 (first entry)

XX

DE Modified Escherichia coli phytase.

XX

KW Phytase; food supplement; enzyme delivery matrix; soybean meal;

KW thermotolerance; thermostability; kernel; phytate;

KW myo-inositol-hexaphosphate; inositol; inorganic phosphate;

KW thermotolerant; feed value; digestion; enzyme.

XX

OS Synthetic.

OS Escherichia coli.

XX

FN US2003103958-A1.

XX

PD 05-JUN-2003.

XX

PF 24-MAY-2002; 2002US-00156660.

XX

PR 13-AUG-1997; 97US-00910798.

PR 01-MAR-1999; 99US-00259214.

PR 13-APR-1999; 99US-00291931.

PR 25-MAY-1999; 99US-00318528.

PR 25-MAY-2000; 2000US-00580515.

PR 24-MAY-2001; 2001US-00866379.

XX

PA (DIVE-) DIVERSA CORP.

XX

PI Short JM, Kretz K, Gray KA, Barton NR, Garrett JB, O'donoghue E;

PI Mathur EU;

XX

DR WPI; 2003-787039/74.  
XX N-PSDB; ADC87742.  
PT New nucleic acid encoding a polypeptide having phytase activity, useful  
PT in improving the feeding value of phytate rich ingredients or as an aid  
XX in phytate digestion.  
XX Claim 60; SEQ ID NO 2; 113pp; English.  
XX  
CC The invention discloses a new isolated or recombinant nucleic acid which  
CC encodes a polypeptide having a phytase activity. Also claimed is a  
CC nucleic acid probe, an amplification primer sequence pair, an expression  
CC cassette comprising the nucleic acid, a vector comprising the nucleic  
CC acid, a transgenic non-human animal or plant, or its seed, comprising the  
CC nucleic acid, an antisense oligonucleotide, inhibiting the translation of  
CC a phytase message in a cell, a heterodimer comprising the polypeptide and  
CC a second domain, an array comprising immobilised polypeptide or nucleic  
CC acid, a hybridoma comprising an antibody that specifically binds to the  
CC polypeptide, a food supplement for an animal, an edible enzyme delivery  
CC matrix, an edible pellet comprising a granule edible carrier and the  
CC polypeptide, a feed composition, a soybean meal, isolating or identifying  
CC the polypeptide, making an anti-phytase antibody, producing a recombinant  
CC polypeptide, determining whether a compound binds to the polypeptide,  
CC identifying a modulator, whole cell engineering of new or modified  
CC phenotypes by using real-time metabolic flux analysis, increasing  
CC the resistance or thermostability of the phytase polypeptide, increasing  
CC the resistance of the phytase polypeptide to enzymatic inactivation in a  
CC digestive system and processing of corn and sorghum kernels. The phytase  
CC activity comprises catalysis of phytate (myo-inositol-hexaphosphate) to  
CC inositol and inorganic phosphate and is thermostable. The nucleic acid  
CC is useful in improving the feeding value of phytate rich ingredients or  
CC as an aid in phytate digestion. The sequence presented is the modified  
CC Escherichia coli phytase.  
XX  
SQ Sequence 436 AA;  
Query Match 43.1%; Score 948.5; DB 7; Length 436;  
Best Local Similarity 48.2%; Pred. No. 1.9e-85;  
Matches 197; Conservative 62; Mismatches 135; Indels 15; Gaps 7;  
QY 13 LVLMSGLAITAPVA---REPSGYTLERVVLSRHGVRPTKQTOLMNDVTPDKKQWP 69  
DB 5 LIPFLSLIPUTPOSAFAQSEPE-LKLESVVIVSRHGVRAPTATQLMQDVTDAWPTWP 63  
QY 70 VKAGYLPRAELVTLMGFGYDFRSIGLL-AAGCPAEGGVYAQADIDORTLTGOAFL 128  
DB 64 VKLGELTPRGELIAYLGHYWRQLVADGLLPKCGCPQSGQVAILADVDERTRKTGEAFA 123  
QY 129 DGVPAGCGLTVHNOADLKKTDPLPHVETGVCKLDNAQTKAIEERLGGPLDVTVSQRYAK 188  
DB 124 AGLAPDCAITVHTQADTSSEPDLENPLKTGVCOLDNANVTDAILERAGGSIADFTGHYQT 183  
QY 189 PPAQMGVNLNFAAPCYCKSLQQCKTKDFAHFAANVNVNKEGKVTLSGLPLALSSTLGE 248  
DB 184 AFRLELVNLPQNCLKREKQDESCLTQALPSELKVSAD--CVSLTGAVSLASMLTE 241  
QY 249 IFLLQNAQAMPEVAWQRLKGAENVVLSLHNAQFNLMAKTPIYARHKGFTPLQQIDTAL 308  
DB 242 IFLLQQAQCMPEPGWGRITDQHWNTLLSHNAQFDLLQRTPEVARSRAETFLDLKNTAL 301  
QY 309 T---LQIDAQOKLPIGAQNEVFLGHDNTNANIAGMLGADWOLPEQDNTPPGGGLVF 365  
DB 302 TPHPQKQAYGVTLPTS-----VLFIAHDTNLMGLGALSNLWTLPCQPNTPPGSELVF 357  
QY 366 ELWQNPNDHORYAVKMYQTMQDLRNAEKLDLKNPAGIISVAVAGE 414  
DB 358 ERWRRLSDNSQWIVQSLVFQTLQWRDKTPLSL-NTPPGVEKLTLAGCE 405  
RESULT 8  
ID ABP51937 standard; protein; 410 AA.  
XX

AC ABP51937;  
XX 08-OCT-2002 (first entry)  
XX Phytase protein sequence.  
DE Phytase; enzyme; E.C.3.1.3.8; phytate; foodstuff; feed.  
XX Unidentified.  
XX WO200248332-A2.  
XX 20-JUN-2002.  
XX 12-DEC-2001; 2001WO-US048774.  
XX 12-DEC-2000; 2000US-0255090P.  
XX (DIVE-) DIVERSA CORP.  
XX Short J, Mathur EJ, Richardson T, Robertson D, Barton N;  
XX WPI; 2002-583504/62.  
XX Novel recombinant phytase protein and polynucleotide for improving  
XX nutritional value of phytate-containing foodstuff, in animal feed and  
XX feed supplements and to degrade excess phytase from environment or  
XX sample.  
XX Disclosure; Fig 9; 208pp; English.  
XX The present invention describes an isolated phytase protein (I). (I) can  
XX be used for improving the nutritional value of a phytate-containing  
XX foodstuff, by contacting the phytate-containing foodstuff with (I), where  
XX the enzyme catalyses the liberation of inorganic phosphate from the  
XX phytate-containing foodstuff, and so improving the nutritive value of the  
XX contacted foodstuff. The liberation of inorganic phosphate occurs prior  
XX to or after the ingestion of phytate-containing foodstuff by a recipient  
XX organism. Nucleotide sequences (II) encoding (I) can be used for  
XX producing an animal feed, by transforming a plant, plant portion or plant  
XX cell with a nucleic acid, expression vector, comprising (II), culturing  
XX the plant, plant portion or plant cell under conditions in which the  
XX phytase protein is expressed and converting the plant, plant portion or  
XX plant cell into a composition suitable for animal feed. The animal is  
XX preferably a monogastric animal or a ruminant. (I) and (II) are useful in  
XX animal feed and feed supplements as well as in treatments to degrade or  
XX remove excess phytate from the environment or a sample. (I) reduces  
XX phytate levels in animal manure and so reduces the phosphate pollution of  
XX the environment. The present sequence represents a phytase from the  
XX present invention  
XX SQ Sequence 410 AA;  
Query Match 42.9%; Score 943; DB 5; Length 410;  
Best Local Similarity 49.0%; Pred. No. 6.2e-85;  
Matches 191; Conservative 60; Mismatches 127; Indels 12; Gaps 6;  
QY 29 ABPSGYTLERVVLSRHGVRPTKQTOLMNDVTPDKKQWPVKAGYLPRAELVTLMGG 88  
DB 2 SEPE-LKLESVVIVSRHGVRAPTATQLMQDVTDAWPTWFKLGELTPRGELIAYLGH 60  
QY 89 FYGDYFRSILGLL-AAGCPAEGGVYAQADIDORTLTGOAFLDGVPAGCGLTVHNOADLKK 147  
DB 61 YWRQLVADGLLPKCGCPQSGQVAILADVDERTRKTGEAFAAGLAPDCAITVHTQADTS 120  
QY 148 TDPLPHVETGVCKLDNAQTKAIEERLGGPLDVTVSQRYAKPFAQMGVNLNFAAPCYCKS 207  
DB 121 PDPLENPLKTGVCOLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLNFPQSNLCLK 180  
QY 208 LQQCKTKDFAHFAANVNVNKEGKVTLSGLPLALSSTLGEIFLLQNAQAMPEVAWQRLK 267  
DB 181 REKQDESCLTQALPSELKVSAD--CVSLTGAVSLASMLTEIFLLQQAQCMPEPGWGRIT 238









64 VKLGWLTFRGGELIAYLGHYQORQLVADGLLAKKGCPSQGVVAIIADVDERTKTGEAFA 123  
 129 DGVAPGCGLTVHNOADLKKTDPLFHPVETGVCKLDNAQTDKATEERLGGPLDTVSQRYAK 188  
 124 AGLAPDCAITVHTQADTSPPDLFNLKTVGCOLDNANVTDAILSRAGSIADFTGHRQT 183  
 189 PFAQMGDVLFNAASPYCKSLQOQKTCDFAHFAANEVNVNKEGTVKLSGLPLALSSTLGE 248  
 184 AFRELRLVNFQSNCLKREKQDESCSLTQALPSELKVSAD--NVSLTGAVSLASMLTE 241  
 249 IFLIQNAQAMPEVAVQRLKGAENWVLSLSLHNAQFNLMAKTPYIARHKGTPLLQOQIDTAL 308  
 242 IFLQQAQMGPEPGWGRITDSHOWNTLSLHNAQFYLLOQTPEVARSRAPELDDLKLTAL 301  
 309 T---LQIDAQOQKLPISAQNRVFLGGHDTNINIANIAGMLGADWQLPEQPDNTPPGGGLVF 365  
 302 TPHPQKQAGVTLPTS---VLFIAGHDTNLANLGALNLTLPQPDNTPPGGGLVF 357  
 366 ELWQNPNDHORYAVKMFYQTMQDLRNAEKLDLKNNPAGIISVAVAGE 414  
 358 ERWRRLSDNSQWIOVSLVFQTLQQRDKTFLSL-NTPPGEVKLTAGCE 405  
  
 RESULT 14  
 AAB36259  
 ID AAB36259 standard; protein; 432 AA.  
 AC AAB36259;  
 XX  
 XX  
 DT 12-SEP-2003 (revised)  
 DT 20-FEB-2001 (first entry)  
 XX  
 DE R15/APPA plasmid translated sequence.  
 XX  
 KW Transgenic animal; salivary protein; phytase; phosphorus; animal growth;  
 KW environmental pollution; pig.  
 OS Rattus sp.  
 OS Escherichia coli.  
 OS Chimeric.  
 FN WO200064247-A1.  
 XX  
 XX 02-NOV-2000.  
 PF 20-APR-2000; 2000WO-CA000430.  
 XX  
 XX 23-APR-1999; 99US-0130508P.  
 XX  
 XX (UYGU-) UNIV GUELPH.  
 PI Forsberg CW, Golovan S, Phillips JP;  
 DR WPI; 2000-687245/67.  
 DR N-PSDB; AAC68296.  
 XX  
 PT Transgenic non-human animal for gastrointestinal tract specific  
 PT expression of a protein, preferably phytase, comprises a nucleic acid  
 PT sequence including a heterologous transgene construct encoding the  
 PT protein.  
 XX  
 PS Disclosure; Fig 19; 152pp; English.  
 XX  
 CC The present invention provides transgenic animals which produce desired  
 CC proteins, in this case pigs which expresses phytase in the salivary  
 CC gland. Low phytase production levels result in phytate in the diet being  
 CC excreted and causing phosphorus contamination in water, as well as  
 CC reducing the growth of animals. The invention provides a number of  
 CC transgenes containing the E. coli APPA phytase coding sequence. (Updated  
 CC on 12-SEP-2003 to standardise OS field)  
 XX  
 SQ Sequence 432 AA;

Query Match 42.7%; Score 939.5; DB 3; Length 432;  
 Best Local Similarity 47.9%; Pred. No. 1.5e-84;  
 Matches 196; Conservative 61; Mismatches 137; Indels 15; Gaps 7;  
  
 QY 13 LVLMISGLAIAITAPVA---AEPSCYTILRVILSRHGVRSPTKOTQLMNDVTPDKWQWP 69  
 DB 5 LIPFLSLIPLTPOSATFAQSEPE-LKLESVVIVSRHGVRAPTKATQLMQDVTDPAMWTP 63  
 QY 70 VKAGVLTFRGAEVLTLMGGFYDYFRSILGIIA-AGCPAEGGVAAQADIDORTRLTGOAFL 128  
 DB 64 VKLGWLTFRGGELIAYLGHYQORQLVADGLLAKKGCPSQGVVAIIADVDERTKTGEAFA 123  
 QY 129 DGVAPGCGLTVHNOADLKKTDPLFHPVETGVCKLDNAQTDKATEERLGGPLDTVSQRYAK 188  
 DB 124 AGLAPDCAITVHTQADTSPPDLFNLKTVGCOLDNANVTDAILSRAGSIADFTGHRQT 183  
 QY 189 PFAQMGDVLFNAASPYCKSLQOQKTCDFAHFAANEVNVNKEGTVKLSGLPLALSSTLGE 248  
 DB 184 AFRELRLVNFQSNCLKREKQDESCSLTQALPSELKVSAD--NVSLTGAVSLASMLTE 241  
 QY 249 IFLIQNAQAMPEVAVQRLKGAENWVLSLSLHNAQFNLMAKTPYIARHKGTPLLQOQIDTAL 308  
 DB 242 IFLQQAQMGPEPGWGRITDSHOWNTLSLHNAQFYLLOQTPEVARSRAPELDDLKLTAL 301  
 QY 309 T---LQIDAQOQKLPISAQNRVFLGGHDTNINIANIAGMLGADWQLPEQPDNTPPGGGLVF 365  
 DB 302 TPHPQKQAGVTLPTS---VLFIAGHDTNLANLGALNLTLPQPDNTPPGGGLVF 357  
 QY 366 ELWQNPNDHORYAVKMFYQTMQDLRNAEKLDLKNNPAGIISVAVAGE 414  
 DB 358 ERWRRLSDNSQWIOVSLVFQTLQQRDKTFLSL-NTPPGEVKLTAGCE 405  
  
 RESULT 15  
 AAB36258  
 ID AAB36258 standard; protein; 432 AA.  
 XX  
 AC AAB36258;  
 XX  
 DT 12-SEP-2003 (revised)  
 DT 20-FEB-2001 (first entry)  
 XX  
 DE R15/APPA plasmid translated sequence.  
 XX  
 KW Transgenic animal; salivary protein; phytase; phosphorus; animal growth;  
 KW environmental pollution; pig.  
 OS Rattus sp.  
 OS Escherichia coli.  
 OS Chimeric.  
 FN WO200064247-A1.  
 XX  
 XX 02-NOV-2000.  
 PF 20-APR-2000; 2000WO-CA000430.  
 XX  
 XX 23-APR-1999; 99US-0130508P.  
 XX  
 XX (UYGU-) UNIV GUELPH.  
 PI Forsberg CW, Golovan S, Phillips JP;  
 DR WPI; 2000-687245/67.  
 DR N-PSDB; AAC68296.  
 XX  
 PT Transgenic non-human animal for gastrointestinal tract specific  
 PT expression of a protein, preferably phytase, comprises a nucleic acid  
 PT sequence including a heterologous transgene construct encoding the  
 PT protein.  
 XX  
 PS Disclosure; Fig 18; 152pp; English.

XX The present invention provides transgenic animals which produce desired  
CC proteins, in this case pigs which expresses phytase in the salivary  
CC gland. Low phytase production levels result in phytate in the diet being  
CC excreted and causing phosphorus contamination in water, as well as  
CC reducing the growth of animals. The invention provides a number of  
CC transgenes containing the E. coli APPA phytase coding sequence. (Updated  
CC on 12-SEP-2003 to standardise CC field)

Sequence 432 AA:

Query Match 42.7%; Score 939.5; DB 3; Length 432;

Best Local Similarity	42.18	66.61	55.53	52.51
Matches 196; Conservative	61	Mismatches 137	Indels 15	Gaps 7

13 LVLMLSLGLAAITAPVA--AEPGTYTLERVILSRHGVRSPTKQTOLMNDVTPDKWPQWP 69

Db

QY 70 VKAGYLTPRGAELVTLMGGFYGDYFRSLGLLA-AGCPAEGGVYAQADIDQTRRLTGQAF 128

D**b** 64 VKLGWLT  
PRGGELIAYLGHYQ  
RQRLVADGLLAKK  
GCPQSGQVAIIAD  
VDERTRTKTGEAFA 123

QY 129 DGVAPGCCGLTVHNQADLKKTDPLFHPVETGVCKLDNAQTDKAIERLGGPLD TVSQRYAK 188

Db 124 AGLAPDCAITVHTQADTSSPDPLFNPLKTGVCQLDNANVTDAILSRAGGSIADFTGHRQT 183

QY 189 PFAQMGDVLFNFAASPYCKSLQQQKTCDFAHFAANEVNVNKEGTVTLGGPLALSSTLGE 248

```
Db      184 AFRELERVLFPPQSNLCLKREKQDESCSLTQALPSELKVSAD--NVSLTGAVSLASMLTE 241
```

249 IFLLQNAQAMPEVAWQRLKGAENWVSLSLHNAQFNLMAKTPYIARHKGTPLLQQIDTAL 308 QY

D5 242 IFLLQQAQMPGPWGGRITDShQWNTLLSLHNAQFYLLQRTPEVARSRATPLDLIKTAL 301

QY 309 T---LQLDAQGQKLPISAQNRVFLGGHDTNIAIAGMLGADWQLEQPDNTPPGGGLVF 365

D**b** 302 T P H P P Q K Q A Y G V T L P T S --- V L F I A G H D T N L A N L G A L E L N W T L P G Q P D N T P P G G E L V F 357

QY 366 ELWQNPDNHQR YVAVKMFYQ TMDQLRNAEKL DLKNNPAGIISVAVAGCE 414

358 ERWRLSDNSQWIQVSLVFQTLQQMRDKTPLSL-NTPPGEVKLTLAGCE 405

Search completed: April 30, 2004, 12:38:45  
Job time : 55.7805 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 12:36:16 ; Search time 16.8293 seconds  
(without alignments)  
1288.405 Million cell updates/sec

Title: US-10-021-723B-4

Perfect score: 2200  
Sequence: 1 MSVLENRVRLSGLVLMISGL.....NPAGIISVAVAGCENNGDDK 420

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/2/iaa/PCUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	932.5	42.4	440	3	US-09-259-214-2
2	932.5	42.4	440	3	US-09-318-528-2
3	932.5	42.4	440	3	US-09-291-931-2
4	928.5	42.2	433	4	US-09-540-149A-1
5	860	39.1	423	2	US-08-910-798-2
6	498	22.6	421	4	US-09-489-039A-7512
7	440.5	20.0	522	4	US-09-489-039A-13501
8	115	5.2	443	4	US-09-684-855-112
9	115	5.2	443	4	US-09-684-855-134
10	115	5.2	443	4	US-09-684-855-157
11	115	5.2	450	4	US-09-044-718-9
12	115	5.2	466	3	US-08-968-435-31
13	115	5.2	466	4	US-08-744-231-31
14	115	5.2	466	4	US-09-273-871A-13
15	99	4.5	2231	4	US-09-252-991A-21854
16	98.5	4.5	1257	4	US-09-336-987A-5
17	97.5	4.4	4928	3	US-09-036-987A-5
18	97.5	4.4	4928	3	US-09-370-700-5
19	97.5	4.4	4928	4	US-09-603-207-5
20	96.5	4.4	947	4	US-09-543-681A-6556
21	94	4.3	449	3	US-08-680-506-7
22	94	4.3	660	4	US-09-107-532A-6715
23	93.5	4.2	504	4	US-09-543-681A-6521
24	93.5	4.2	1584	4	US-09-457-040B-27
25	92.5	4.2	439	4	US-09-543-681A-7293
26	91.5	4.2	394	6	5223418-2
27	91.5	4.2	529	1	US-08-548-509-2

Query Match		42.4%	Score 932.5;	DB 3;	Length 440;
Best Local Similarity		47.7%	Pred. No. 4.9e-92;		
Matches		195;	Conservative	61;	Mismatches 138;
				Indels	15;
				Gaps	7;
QY	13	LVLMLSLAIAITAPVA---AEPGYTLERVVILSRHGVRSPTQTQTLMDNVTDPKWPWP	69		
DB	5	LIPFLSLILPLTPQSAFAQSEPE-LKLESVIVSRHGVRAPTKATQLMQDVTDPDAMPWP	63		
QY	70	VKAGYLTPRGAELVTLMGFGDYFRSLGLLA-AGCPAEGGVAAQADIDORTLTGQAF	128		
DB	64	VKLGLTPRGELIAYIGHYQRLVADGLLAKKGCPSQGVAIADVDERTKTGEAFA	123		
QY	129	DGVAPGCLTVHNOADLKKTDPLPHVETGVCKLDNAQTDKAISERLGGPLDVTVSQYAK	189		
DB	124	AGLAPDCAITVHTQADTSSPDPLFNPLKTVGCQLDANAVTDAILSRAGGSIAIDFTGHRT	183		
QY	189	PFAQMGDVIAFAAPSPYCKSLQOQKTCDFAHFAANVNVNKEGTVKLTSGPLALSSTLGE	248		
DB	184	AFRESLRLVFPQSNLCIKREKQDSCSLTQALPSELKVSAD--NVSLTGAVSLSMLTE	241		
QY	249	IFLLQNAQAPPEVAWQRLKGAENWVSLSLHNAQFNLMAKTPYIARHKGTPLLQCIDTAL	308		
DB	242	IFLLQNAQAPPEVAWQRLKGAENWVSLSLHNAQFNLMAKTPYIARHKGTPLLQCIDTAL	301		
QY	309	T---LQIDAQOQKLPISAQNRVFLPGHDTNIAIAGMIGADWOLPQDPONTPEGGGLVF	365		
DB	302	TPHPQKQAYGVTLPTS-----VLFIAGHDTNLNLGALLENMTLPQDPONTPEGGGLVF	357		
QY	366	ELWQNPQNHQRYAVKMFYQTMQDLRQNAEKLDLKNPAGIISVAVAGCE	414		

ALIGNMENTS

RESULT 1  
US-09-259-214-2  
; Sequence 2, Application US/09259214A  
; Patent No. 6110719  
; GENERAL INFORMATION:  
; APPLICANT: Kretz, Keith  
; TITLE OF INVENTION: NOVEL PHYTASE  
; FILE REFERENCE: DIVER1370-1  
; CURRENT APPLICATION NUMBER: US/09/259,214A  
; CURRENT FILING DATE: 1999-03-01  
; EARLIER APPLICATION NUMBER: 08/910,798  
; EARLIER FILING DATE: 1997-08-13  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 440  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-259-214-2

Query Match 42.4%; Score 932.5; DB 3; Length 440;  
Best Local Similarity 47.7%; Pred. No. 4.9e-92;  
Matches 195; Conservative 61; Mismatches 138; Indels 15; Gaps 7;

Db 358 ERWRRLSDNSQWVQVSLVFTLQMQMDKTPSL-NTPPGEVKLTLAGCE 405

RESULT 2

US-09-318-528-2  
; Sequence 2, Application US/09318528  
; Patent No. 6183740  
; GENERAL INFORMATION:  
; APPLICANT: Kretz, Keith  
; TITLE OF INVENTION: NOVEL PHYTASE  
; FILE REFERENCE: 09010/029003  
; CURRENT APPLICATION NUMBER: US/09/318,528  
; CURRENT FILING DATE: 1999-05-25  
; EARLIER APPLICATION NUMBER: 09/291,931  
; EARLIER FILING DATE: 1999-04-13  
; EARLIER APPLICATION NUMBER: 08/910,798  
; EARLIER FILING DATE: 1997-08-13  
; EARLIER APPLICATION NUMBER: 09/259,214  
; EARLIER FILING DATE: 1999-03-01  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 440  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-318-528-2

Query Match 42.4%; Score 932.5; DB 3; Length 440;  
Best Local Similarity 47.7%; Pred. No. 4.9e-92;  
Matches 195; Conservative 61; Mismatches 138; Indels 15; Gaps 7;  
QY 13 LVLMGLAAITAPVA---AEPGTYLTVVILSRHGVRSPTKOTQLMNDVTPDKWPMP 69  
DB 5 LIPFLSLIPLTPQSAFAQSEPE-LKLESVVIVSRHGVRAPTKATQLMQDVTDPANWTP 63  
QY 70 VKAGYLPTRGAEVTLMGFGYDYFRSLGLLA-AGCPAEGGVYAQADIDORTLTGOAFL 128  
DB 64 VLGWLTPRGELTAYLGHYQORLVADGLLAKKGCPSQGVAILADVERTKTKGEAFA 123  
QY 129 DGVAPEGGLTVHQAADLKTDPLPHVETGVCKLDNAQTDKAIEERLGGPLDVTVSRYAK 188  
DB 124 AGLAPDCAITVHTQADTSSPDLPFNPLKTVGCQLDNANVTDAILSRAGGSIAFTGHRQT 183  
QY 189 PFAQMGDVLPFAASPYCKSLQOQKTCDFAHFAANEVNVNKEGTVKLTSGPLALSSTLGE 248  
DB 184 AFRELERVLPQSNLCKREKQDESCSLTQALPSELKVSAD--NVSLTGAVSLASMLTE 241  
QY 249 IFLQNAQAMPEVAVQRLKGAENVVSLSLHNAQFNLMAKTPYIARHKGTPLLQQIDTAL 308  
DB 242 IFLQQAQGMPEPGWGRITDSHQNTLLSLHNAQFYLLQRTPEVARSRATPDLDLMAAL 301  
QY 309 T---LQDAQOQKLPISAQNRVFLGHDNTNIAAGMLGADWQLPEQPDNTPPGGGLVF 365  
DB 302 TPHPPOQAQYGVTLPTS---VLFIAHDNTNLANLGGALELNTLPGQPDNTPPGGELVF 357  
QY 366 ELNQNPNDHQRVYAVKMFYQTMQDLRNAEKLDKNPAAGIISVAVAGCE 414  
DB 358 ERWRRLSDNSQWVQVSLVFTLQMQMDKTPSL-NTPPGEVKLTLAGCE 405

RESULT 3

US-09-291-931-2  
; Sequence 2, Application US/09291931A  
; Patent No. 6190897  
; GENERAL INFORMATION:  
; APPLICANT: Kretz, Keith  
; TITLE OF INVENTION: NOVEL PHYTASE  
; FILE REFERENCE: 09010/029003  
; CURRENT APPLICATION NUMBER: US/09/291,931A  
; CURRENT FILING DATE: 1999-04-13  
; EARLIER APPLICATION NUMBER: 08/910,798  
; EARLIER FILING DATE: 1997-08-13  
; EARLIER APPLICATION NUMBER: 09/259,214

; EARLIER FILING DATE: 1999-03-01  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 440  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-291-931-2

Query Match 42.4%; Score 932.5; DB 3; Length 440;  
Best Local Similarity 47.7%; Pred. No. 4.9e-92;  
Matches 195; Conservative 61; Mismatches 138; Indels 15; Gaps 7;  
QY 13 LVLMGLAAITAPVA---AEPGTYLTVVILSRHGVRSPTKOTQLMNDVTPDKWPMP 69  
DB 5 LIPFLSLIPLTPQSAFAQSEPE-LKLESVVIVSRHGVRAPTKATQLMQDVTDPANWTP 63  
QY 70 VKAGYLPTRGAEVTLMGFGYDYFRSLGLLA-AGCPAEGGVYAQADIDORTLTGOAFL 128  
DB 64 VLGWLTPRGELTAYLGHYQORLVADGLLAKKGCPSQGVAILADVERTKTKGEAFA 123  
QY 129 DGVAPEGGLTVHQAADLKTDPLPHVETGVCKLDNAQTDKAIEERLGGPLDVTVSRYAK 188  
DB 124 AGLAPDCAITVHTQADTSSPDLPFNPLKTVGCQLDNANVTDAILSRAGGSIAFTGHRQT 183  
QY 189 PFAQMGDVLPFAASPYCKSLQOQKTCDFAHFAANEVNVNKEGTVKLTSGPLALSSTLGE 248  
DB 184 AFRELERVLPQSNLCKREKQDESCSLTQALPSELKVSAD--NVSLTGAVSLASMLTE 241  
QY 249 IFLQNAQAMPEVAVQRLKGAENVVSLSLHNAQFNLMAKTPYIARHKGTPLLQQIDTAL 308  
DB 242 IFLQQAQGMPEPGWGRITDSHQNTLLSLHNAQFYLLQRTPEVARSRATPDLDLMAAL 301  
QY 309 T---LQDAQOQKLPISAQNRVFLGHDNTNIAAGMLGADWQLPEQPDNTPPGGGLVF 365  
DB 302 TPHPPOQAQYGVTLPTS---VLFIAHDNTNLANLGGALELNTLPGQPDNTPPGGELVF 357  
QY 366 ELNQNPNDHQRVYAVKMFYQTMQDLRNAEKLDKNPAAGIISVAVAGCE 414  
DB 358 ERWRRLSDNSQWVQVSLVFTLQMQMDKTPSL-NTPPGEVKLTLAGCE 405

RESULT 4

US-09-540-149A-1  
; Sequence 1, Application US/09540149A  
; Patent No. 6511699  
; GENERAL INFORMATION:  
; APPLICANT: Lei, Xingen  
; TITLE OF INVENTION: ENZYMES WITH IMPROVED PHYTASE ACTIVITY  
; FILE REFERENCE: 19603/2791  
; CURRENT APPLICATION NUMBER: US/09/540,149A  
; CURRENT FILING DATE: 2000-03-31  
; PRIOR APPLICATION NUMBER: 60/127,032  
; PRIOR FILING DATE: 1999-03-31  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 433  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (433)  
; OTHER INFORMATION: Xaa at position 433 in this sequence is unknown  
US-09-540-149A-1

Query Match 42.2%; Score 928.5; DB 4; Length 433;  
Best Local Similarity 47.4%; Pred. No. 1.3e-91;  
Matches 194; Conservative 63; Mismatches 137; Indels 15; Gaps 7;  
QY 13 LVLMGLAAITAPVA---AEPGTYLTVVILSRHGVRSPTKOTQLMNDVTPDKWPMP 69  
DB 5 LIPFLSLIPLTPQSAFAQSEPE-LKLESVVIVSRHGVRAPTKATQLMQDVTDPANWTP 63

QY 70 VKAGYLTFRGAEVLTLMGGFYGDYFRSGLLIA-ACCPAEGGVYAAQADIDORTLTGQAF 128  
DB 64 VKLWLTFRGAEVLTLMGGFYGDYFRSGLLIA-ACCPAEGGVYAAQADIDORTLTGQAF 119  
QY 129 DGVAPGCGLTVMNQADLKTDPFHVETGVCKLDNAQTDKAIERLGGPLDVTVSQRYAK 188  
DB 124 AGLAPDCAITVHTQADTSSPDLFNPVKTGVQCLDNANVTDAISRAGGSIAFTGHRQT 166  
QY 189 PFAQMGDVINFAASPYCKSLQOQKTCDFAHFAANVNNKGTGKVTLSGPLALSSTIGE 248  
DB 184 AFRELERVNLFPQSLCLNREKQDESCSLTQALPSELKVSAD--NVSLTGAVSLASMLTE 224  
QY 249 IFLLQNAQAMPVAVORLKGAEVWVLSLHNAQFNLMAKTPYIARHKGTPLLQOQIDTAL 308  
DB 242 IFLLQNAQAMPVAVORLKGAEVWVLSLHNAQFNLMAKTPYIARHKGTPLLQOQIDTAL 284  
QY 309 T---LQDAAQOQKLPISAQNRVFLFGGHDNTIANIAGMLGADWQOLPEQDNTPPGGGLVF 365  
DB 302 TPHPQKQAYGVTLPTS---VLFAGHDNTLANLGALELNLWTLFQPDNTPPGGGLVF 340  
QY 366 ELWQNPNDHQRVAVKMFYQTMDOQLRNAEKLDLKNPAGIISVAVAGE 414  
DB 358 ERMRRLSDNSQWISVLFQTLQOQMDKTPLSL-NTPPGEVKLTLAGCE 388

RESULT 5

US-08-910-798-2  
; Sequence 2, Application US/08910798  
; Patent No. 5876997  
; GENERAL INFORMATION:  
; APPLICANT: KRETZ  
; TITLE OF INVENTION: NOVEL PHYTASE  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: California  
; COUNTRY: US  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/910,798  
; FILING DATE: August 13, 1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: HAILE, PH.D., LISA A.  
; REGISTRATION NUMBER: 38,347  
; REFERENCE/DOCKET NUMBER: 09010/029001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619/678-5070  
; TELEFAX: 619/678-5099  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 423 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-910-798-2

Query Match 39.1%; Score 860; DB 2; Length 423;  
Best Local Similarity 45.5%; Pred. No. 3.3e-84;  
Matches 186; Conservative 58; Mismatches 133; Indels 32; Gaps 8;  
QY 13 LVLMGLAAITAPVA---APPSTYLERVILSHGVRSPTKQTLMDVTPDKWPWP 69  
DB 5 LIPFLSLIPLTPQAFQAQSEPE-LKLESVIVSRHVRAPTKATQLMQDVTDAWPTW 63

QY 70 VKAGYLTFRGAEVLTLMGGFYGDYFRSGLLIA-ACCPAEGGVYAAQADIDORTLTGQAF 128  
DB 64 VKLWLTFRGAEVLTLMGGFYGDYFRSGLLIA-ACCPAEGGVYAAQADIDORTLTGQAF 119  
QY 129 DGVAPGCGLTVMNQADLKTDPFHVETGVCKLDNAQTDKAIERLGGPLDVTVSQRYAK 188  
DB 120 -----QADTSSPDLFNPVKTGVQCLDNANVTDAISRAGGSIAFTGHRQT 166  
QY 189 PFAQMGDVINFAASPYCKSLQOQKTCDFAHFAANVNNKGTGKVTLSGPLALSSTIGE 248  
DB 167 AFRELERVNLFPQSLCLNREKQDESCSLTQALPSELKVSAD--NVSLTGAVSLASMLTE 224  
QY 249 IFLLQNAQAMPVAVORLKGAEVWVLSLHNAQFNLMAKTPYIARHKGTPLLQOQIDTAL 308  
DB 225 IFLLQNAQAMPVAVORLKGAEVWVLSLHNAQFNLMAKTPYIARHKGTPLLQOQIDTAL 284  
QY 309 T---LQDAAQOQKLPISAQNRVFLFGGHDNTIANIAGMLGADWQOLPEQDNTPPGGGLVF 365  
DB 285 TPHPQKQAYGVTLPTS---VLFAGHDNTLANLGALELNLWTLFQPDNTPPGGGLVF 340  
QY 366 ELWQNPNDHQRVAVKMFYQTMDOQLRNAEKLDLKNPAGIISVAVAGE 414  
DB 341 ERMRRLSDNSQWISVLFQTLQOQMDKTPLSL-NTPPGEVKLTLAGCE 388

RESULT 6

US-09-489-039A-7512  
; Sequence 7512, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 7512  
; LENGTH: 421  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-7512

Query Match 22.6%; Score 498; DB 4; Length 421;  
Best Local Similarity 30.5%; Pred. No. 5.5e-45;  
Matches 126; Conservative 84; Mismatches 165; Indels 38; Gaps 12;  
QY 10 LSGVLMLSLGLAAITAPVAAPSGYTLERVILSHGVRSPTKQTLMDVTPDKWPWP 68  
DB 16 VTGAVMLSG---AQADKAAPEGYLOQVLMRSHRNLRAPLANNGLVLEQSTAKWPWP 72  
QY 69 PVKAGYLTFRGAEVLTLMGGFYGDYFRSGLLIAAG-CPAEGGVYAAQADIDORTLTGQAF 127  
DB 73 DVPGGQLTGKGVLEVYMGHVMREMLAQOKLVTSGECPENAVAVANSQRTVATAQPF 132  
QY 128 LDGVAPGCGLTVMNQADLKTDPFHVETGVCKLDNAQTDKAIERLGGPLDVTVSQRYA 187  
DB 133 ITGAFPGCGIPVHHQPMQMGTMDFTFNPVITDPSAPFAKALQAMKEROG-----MQT 186  
QY 188 KPFAQMGDVINFAASPYCKSLQOQKTCDFAH---FAANVNNKGTGKVTLSGPLALS 243  
DB 187 ESKLLETWIDYNSPSCK---EKKVLSLSEKGTTFSSAGYQO-----EPGVSGLPKVG 236  
QY 244 STIGEITPLQNAQAMP--EVAVORLKGAEVWVLSLHNAQFNLMAKTPYIARHKGTPLL 301  
DB 237 NSLVDAFTLQYVEGFPKQVAMGEIASDKQWRVLSKLNKNGYQDSLFTSVAVAQNAKPLV 296  
QY 302 QQIDTALTQLDAQOQKLPISAQNRVFLFGGHDNTIANIAGMLG-ADWOLPEQDNTPPG 360  
DB 297 KYIDNALV----GEG-----ASKAKVTLVGHDSNTIASLTALDFXPYQLPQYETPIG 347



QY 361 GGLVFELWQNDHQRVAVKMFYQTMQDLRNEAKLDLKNPAGIISVAVAGC 413  
Db 348 GKLLFQWHDSDAGNRDLKIEYYQSTEQRLRNADALTQAPQR-VTIALNGC 399

RESULT 7  
US-09-489-039A-13501  
; Sequence 13501, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; FILE REFERENCE: 2709.2004001  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR FILING DATE: 2000-01-27  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 13501  
; LENGTH: 522  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-13501

Query Match 20.0%; Score 440.5; DB 4; Length 522;  
Best Local Similarity 31.7%; Pred. No. 1.4e-38;  
Matches 126; Conservative 62; Mismatches 178; Indels 31; Gaps 13;  
QY 9 RLGLV-LMLSLAAITAPVAAPSGYTLERVVILSRHGVRSPTK-OTOLMNDVTPDKWP 66  
Db 105 RHQGLRLFTACALPLLALQSAADWQLEKVVLSRHGIRPTAGNREAZAAGRPT 164  
QY 67 QWPKAGYLTPRGAEVTLMGFGYDFRSLGLLAAGCPAEGGVYAQAIDORTRLTGOA 126  
Db 165 EMTDHGELTGHGVAAYVKNRGAEGOHYQLGLQAGCPAEBIYVRASFLQRTATAQA 224  
QY 127 FLQGVAPCGGLTVHNOADLKKTDPLFHPVETGVCKLDNAQTKAIEERLGGPLDVSQRY 186  
Db 225 LVDAFPGCGVAIHVYSG--DADPLFTQDKFAOTDTPARQLAAVKEAG--DLAQRQ 279  
QY 187 AKPFAQMGVNLFAAPSYCKSLQOQKTCDFAHFAANEVNVNKEGKYVLSPLALSSLT 246  
Db 280 A-----LAPTIQLLKQAVC---QADKPCPIFD-TPWQVEQSKG-KTTISG-LSVMANN 327  
QY 247 GEIFLLQNAAMP--EVAMORLKAENWVLSLHNAQNLMAKTPYIARHKGTPLLQOI 304  
Db 328 VETRLQWSENPLSQLANGKITCARQITALLPTEYDITNDVLYTAQKRGSVLLNAM 387  
QY 305 DTALTQLDAGQKLPISQNRVFLFGGHDNTIANIAGMLGADWQLP-BOPDNTPPGGL 363  
Db 388 -----LD--GVKPEANPNVRWLLLVADHTNIAMVRLTMNFSWQLPGYSRGNIPPGSL 438  
QY 364 VFELWQNDHQRVAVKMFYQTMQDLRNEAKLDLKN 400  
Db 439 VLERWNAKGERVLYRFOAGLDDLRRLQTTDAQH 475

RESULT 8  
US-09-684-855-112  
; Sequence 112, Application US/09684855  
; Patent No. 6599735  
; GENERAL INFORMATION:  
; APPLICANT: F. Hoffmann-La Roche AG  
; TITLE OF INVENTION: CONTINUOUS FERMENTATION PROCESS  
; FILE REFERENCE: C38435/111692  
; CURRENT FILING DATE: 2000-10-06  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR FILING DATE: 1999-10-11  
; NUMBER OF SEQ ID NOS: 169  
; SEQ ID NO 112  
; LENGTH: 443  
; TYPE: PRT  
; ORGANISM: T. thermophilus  
US-09-684-855-112

Query Match 5.2%; Score 115; DB 4; Length 443;  
Best Local Similarity 21.1%; Pred. No. 0.0019;  
Matches 87; Conservative 53; Mismatches 134; Indels 118; Gaps 18;  
QY 31 PSGYTLERVVILSRHGVRSPTKQ-----TOLMNDVTPD-----KWPQWPKVAGY 74  
Db 42 PQNCKITFVQLLSRHGARYPTSSKTELYSQLISRIQKTATAYKGYAFLKDYRYQLGAND 101  
QY 75 LTPRGAELVTLMGFGYDFRSLGLLAAGCPAEGGVYAQAIDORTRLTGOAFLDGVAPG 134  
Db 75 LTPRGAELVTLMGFGYDFRSLGLLAAGCPAEGGVYAQAIDORTRLTGOAFLDGVAPG 134  
QY 102 LTPFGENQMIQIGIKFYNYHYSKLARNV-----PFVRCGSDRVIASGRFLFIEGQSA 154  
Db 102 LTPFGENQMIQIGIKFYNYHYSKLARNV-----PFVRCGSDRVIASGRFLFIEGQSA 154

RESULT 9  
US-09-684-855-134  
; Sequence 134, Application US/09684855  
; Patent No. 6599735  
; GENERAL INFORMATION:  
; APPLICANT: F. Hoffmann-La Roche AG  
; TITLE OF INVENTION: CONTINUOUS FERMENTATION PROCESS  
; FILE REFERENCE: C38435/111692  
; CURRENT FILING DATE: 2000-10-06  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR FILING DATE: 1999-10-11  
; NUMBER OF SEQ ID NOS: 169  
; SEQ ID NO 134  
; LENGTH: 443  
; TYPE: PRT  
; ORGANISM: T. thermophilus  
US-09-684-855-134

Query Match 5.2%; Score 115; DB 4; Length 443;  
Best Local Similarity 21.1%; Pred. No. 0.0019;  
Matches 87; Conservative 53; Mismatches 134; Indels 118; Gaps 18;  
QY 31 PSGYTLERVVILSRHGVRSPTKQ-----TOLMNDVTPD-----KWPQWPKVAGY 74  
Db 42 PQNCKITFVQLLSRHGARYPTSSKTELYSQLISRIQKTATAYKGYAFLKDYRYQLGAND 101  
QY 75 LTPRGAELVTLMGFGYDFRSLGLLAAGCPAEGGVYAQAIDORTRLTGOAFLDGVAPG 134  
Db 75 LTPRGAELVTLMGFGYDFRSLGLLAAGCPAEGGVYAQAIDORTRLTGOAFLDGVAPG 134  
QY 102 LTPFGENQMIQIGIKFYNYHYSKLARNV-----PFVRCGSDRVIASGRFLFIEGQSA 154  
Db 102 LTPFGENQMIQIGIKFYNYHYSKLARNV-----PFVRCGSDRVIASGRFLFIEGQSA 154

; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 112  
; LENGTH: 443  
; TYPE: PRT  
; ORGANISM: T. thermophilus  
US-09-684-855-112

Query Match 5.2%; Score 115; DB 4; Length 443;  
Best Local Similarity 21.1%; Pred. No. 0.0019;  
Matches 87; Conservative 53; Mismatches 134; Indels 118; Gaps 18;  
QY 31 PSGYTLERVVILSRHGVRSPTKQ-----TOLMNDVTPD-----KWPQWPKVAGY 74  
Db 42 PQNCKITFVQLLSRHGARYPTSSKTELYSQLISRIQKTATAYKGYAFLKDYRYQLGAND 101  
QY 75 LTPRGAELVTLMGFGYDFRSLGLLAAGCPAEGGVYAQAIDORTRLTGOAFLDGVAPG 134  
Db 102 LTPFGENQMIQIGIKFYNYHYSKLARNV-----PFVRCGSDRVIASGRFLFIEGQSA 154  
QY 135 CGLTVHNOADLKKTDPLFHPV-----ETGVCKLDNAQTKAIEERLGGPLDVS 183  
Db 155 KVLDPH--SDKHDAPPTINVIIIEGSPYNNLTGSCFV-----FEDSSG--HDAQ 202  
QY 184 QRYAKPFAQM-----GDVLNFAAPSYCKSLQOQKTCDFAHFAANEVNVNKEGTK 233  
Db 203 EKFAQFAPALEKIKDHLPGVDLAVSDVPLMDL-----CPETLARNHTD----- 249  
QY 234 VTLGSPALSSLTGEIFLLQNAQMPAVQORLKAENWVLSLSLH--NAQFNLMAK--- 288  
Db 250 -TLSFFCALST-----QSEWQ-----AYDYQSLGKYYGNGGNGPLGPAQG 289  
QY 289 ---TPYIARHKGTPLLQOQIDTALTQLDAGQKLPISQNRVFLFGGHDNTIANIAGML 344  
Db 290 VGFVNEILLARHTHSPV--QDYTTVNTLDSNPATFPLNATLYADF--SHDNTWTSIFAAL 345  
QY 345 Q-----ADWOLPEQPD-----NTPFGGLVPELMQNDPNHORYVAV 380  
Db 346 GLYNGTAKLSTTEKSIETDGSAAWTVFGGRAYIEMMCDDSDPEVVRV 397

RESULT 9  
US-09-684-855-134  
; Sequence 134, Application US/09684855  
; Patent No. 6599735  
; GENERAL INFORMATION:  
; APPLICANT: F. Hoffmann-La Roche AG  
; TITLE OF INVENTION: CONTINUOUS FERMENTATION PROCESS  
; FILE REFERENCE: C38435/111692  
; CURRENT FILING DATE: 2000-10-06  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR FILING DATE: 1999-10-11  
; NUMBER OF SEQ ID NOS: 169  
; SEQ ID NO 134  
; LENGTH: 443  
; TYPE: PRT  
; ORGANISM: T. thermophilus  
US-09-684-855-134

Query Match 5.2%; Score 115; DB 4; Length 443;  
Best Local Similarity 21.1%; Pred. No. 0.0019;  
Matches 87; Conservative 53; Mismatches 134; Indels 118; Gaps 18;  
QY 31 PSGYTLERVVILSRHGVRSPTKQ-----TOLMNDVTPD-----KWPQWPKVAGY 74  
Db 42 PQNCKITFVQLLSRHGARYPTSSKTELYSQLISRIQKTATAYKGYAFLKDYRYQLGAND 101  
QY 75 LTPRGAELVTLMGFGYDFRSLGLLAAGCPAEGGVYAQAIDORTRLTGOAFLDGVAPG 134  
Db 75 LTPRGAELVTLMGFGYDFRSLGLLAAGCPAEGGVYAQAIDORTRLTGOAFLDGVAPG 134  
QY 102 LTPFGENQMIQIGIKFYNYHYSKLARNV-----PFVRCGSDRVIASGRFLFIEGQSA 154  
Db 102 LTPFGENQMIQIGIKFYNYHYSKLARNV-----PFVRCGSDRVIASGRFLFIEGQSA 154

QY 135 CGLTVHNOADLKKTDPLFHPV-----ETGVCKLDNAQTDKAIEERLGGPLDTS 183  
Db 155 KVLDPH--SDKHDAPPTINVIIEGSPSYNTLDTGSCP-----FEDSSGG--HDAQ 202  
QY 184 QRYAKPFAQM-----GVLNFAASPYCKSLQOQKTCDFAHFAANVNNKEGTYK 233  
Db 203 EKFAKQFAPALEKIKDHLPGVDLAVSDVPYLMDL-----CPFETLARNHTD----- 249  
QY 234 VTLSPGALSSITLGEIFLLQNAQAMPEVAKORLKGAEWVSLSLH--NAQFNLMAK--- 288  
Db 250 -TLPSPFCALST-----QEWQ-----AYDYOSLGKYGNGGNGPLGPAQ 289  
QY 289 ----TPYIARHKGTPLLQOQIDTALTQLDAQOQKLPISAQNRVFLGGHDTNIAIAGML 344  
Db 290 VGFVNELIARMTHTSPV--QDYTTVNHTLDSNPATFPLNATLYADF--SHDNTMTSIFAAL 345  
QY 345 G-----ADWQLPEQPD-----NTPPGGLVFLWQPNHQRVAV 380  
Db 346 GLYNGTAKLSTTEIKSIEETDGYSAANTVPFGGRAYIEMMQCDDSDPVPVRV 397

RESULT 10  
US-09-684-855-157  
; Sequence 157, Application US/09684855  
; Patent No. 6599735  
; GENERAL INFORMATION:  
; APPLICANT: F. Hoffmann-La Roche AG  
; TITLE OF INVENTION: CONTINUOUS FERMENTATION PROCESS  
; FILE REFERENCE: C38435/111692  
; CURRENT APPLICATION NUMBER: US/09/684,855  
; CURRENT FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: EP 00121663.9  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: EP 99120289.6  
; PRIOR FILING DATE: 1999-10-11  
; NUMBER OF SEQ ID NOS: 169  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 157  
; LENGTH: 443  
; TYPE: PRT  
; ORGANISM: T. thermophilus  
US-09-684-855-157

Query Match 5.2%; Score 115; DB 4; Length 443;  
Best Local Similarity 21.1%; Pred. No. 0.0019;  
Matches 87; Conservative 53; Mismatches 154; Indels 118; Gaps 18;  
QY 31 PSGYTLERVILSRHGVRSPTKQ-----TQLMNDVTPD-----KWPQWPKAGY 74  
Db 42 PQNCKITFVQLLSRGARYPTSSKTLYSLISRIQKTATAYKGYAFKDYRYQLGAND 101  
QY 75 LTPRGAELVTLMGGFYGYFRSLGALLAACGPAEGVYAQADIDORTLTGQAFDGVAPG 134  
Db 102 LTPFGENQMIQLGIKFYNYHKSILARNV-----PFVRCSGSDRVIASGRFLTEGFQSA 154  
QY 135 CGLTVHNOADLKKTDPLFHPV-----ETGVCKLDNAQTDKAIEERLGGPLDTS 183  
Db 155 KVLDPH--SDKHDAPPTINVIIEGSPSYNTLDTGSCP-----FEDSSGG--HDAQ 202  
QY 184 QRYAKPFAQM-----GVLNFAASPYCKSLQOQKTCDFAHFAANVNNKEGTYK 233  
Db 203 EKFAKQFAPALEKIKDHLPGVDLAVSDVPYLMDL-----CPFETLARNHTD----- 249  
QY 234 VTLSPGALSSITLGEIFLLQNAQAMPEVAKORLKGAEWVSLSLH--NAQFNLMAK--- 288  
Db 250 -TLPSPFCALST-----QEWQ-----AYDYOSLGKYGNGGNGPLGPAQ 289  
QY 289 ----TPYIARHKGTPLLQOQIDTALTQLDAQOQKLPISAQNRVFLGGHDTNIAIAGML 344  
Db 290 VGFVNELIARMTHTSPV--QDYTTVNHTLDSNPATFPLNATLYADF--SHDNTMTSIFAAL 345  
QY 345 G-----ADWQLPEQPD-----NTPPGGLVFLWQPNHQRVAV 380

Db 346 GLYNGTAKLSTTEIKSIEETDGYSAANTVPFGGRAYIEMMQCDDSDPVPVRV 397

RESULT 11  
US-09-044-718-9  
; Sequence 9, Application US/09044718  
; Patent No. 6391605  
; GENERAL INFORMATION:  
; APPLICANT: KOSTREWA, Dirk  
; APPLICANT: PASANMONTES, Luis  
; APPLICANT: TOMSCHY, Andrea  
; APPLICANT: van LOON, Adolphus  
; APPLICANT: VOGEL, Kurt  
; APPLICANT: WYSS, Markus  
; TITLE OF INVENTION: MODIFIED PHYTASES  
; FILE REFERENCE: Modified Phytases  
; CURRENT APPLICATION NUMBER: US/09/044,718  
; CURRENT FILING DATE: 1998-03-19  
; PRIOR APPLICATION NUMBER: EP 97810175.6  
; PRIOR FILING DATE: 1997-03-25  
; NUMBER OF SEQ ID NOS: 82  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 9  
; LENGTH: 450  
; TYPE: PRT  
; ORGANISM: Talaromyces thermophilus  
US-09-044-718-9

Query Match 5.2%; Score 115; DB 4; Length 450;  
Best Local Similarity 21.1%; Pred. No. 0.002;  
Matches 87; Conservative 53; Mismatches 154; Indels 118; Gaps 18;  
QY 31 PSGYTLERVILSRHGVRSPTKQ-----TQLMNDVTPD-----KWPQWPKAGY 74  
Db 49 PQNCKITFVQLLSRGARYPTSSKTLYSLISRIQKTATAYKGYAFKDYRYQLGAND 108  
QY 75 LTPRGAELVTLMGGFYGYFRSLGALLAACGPAEGVYAQADIDORTLTGQAFDGVAPG 134  
Db 109 LTPFGENQMIQLGIKFYNYHKSILARNV-----PFVRCSGSDRVIASGRFLTEGFQSA 161  
QY 135 CGLTVHNOADLKKTDPLFHPV-----ETGVCKLDNAQTDKAIEERLGGPLDTS 183  
Db 162 KVLDPH--SDKHDAPPTINVIIEGSPSYNTLDTGSCP-----FEDSSGG--HDAQ 209  
QY 184 QRYAKPFAQM-----GVLNFAASPYCKSLQOQKTCDFAHFAANVNNKEGTYK 233  
Db 210 EKFAKQFAPALEKIKDHLPGVDLAVSDVPYLMDL-----CPFETLARNHTD----- 256  
QY 234 VTLSPGALSSITLGEIFLLQNAQAMPEVAKORLKGAEWVSLSLH--NAQFNLMAK--- 288  
Db 257 -TLPSPFCALST-----QEWQ-----AYDYOSLGKYGNGGNGPLGPAQ 296  
QY 289 ----TPYIARHKGTPLLQOQIDTALTQLDAQOQKLPISAQNRVFLGGHDTNIAIAGML 344  
Db 297 VGFVNELIARMTHTSPV--QDYTTVNHTLDSNPATFPLNATLYADF--SHDNTMTSIFAAL 352  
QY 345 G-----ADWQLPEQPD-----NTPPGGLVFLWQPNHQRVAV 380  
Db 353 GLYNGTAKLSTTEIKSIEETDGYSAANTVPFGGRAYIEMMQCDDSDPVPVRV 404

RESULT 12  
US-08-868-435-31  
; Sequence 31, Application US/08868435  
; Patent No. 6291221  
; GENERAL INFORMATION:  
; APPLICANT: Van Loon, Adolphus  
; APPLICANT: Mitchell, David  
; TITLE OF INVENTION: POLYPEPTIDES WITH PHYTASE ACTIVITY  
; NUMBER OF SEQUENCES: 35  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hoffmann-La Roche Inc.

STREET: 340 Kingsland Street  
 CITY: Nutley  
 STATE: New Jersey  
 COUNTRY: United States of America  
 ZIP: 07110  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent in Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 FILING DATE: US/08/868,435  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/744,231  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Kass, Alan P  
 REGISTRATION NUMBER: 32142  
 REFERENCE/DOCKET NUMBER: Case Docket 9339  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (201) 235-2363  
 TELEFAX: (201) 235-2363  
 INFORMATION FOR SEQ ID NO: 31:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 466 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: 204  
 OTHER INFORMATION: /note="potential N-glycosylation site"  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: 269  
 OTHER INFORMATION: /note="potential N-glycosylation site"  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: 335  
 OTHER INFORMATION: /note="potential N-glycosylation site"  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: 348  
 OTHER INFORMATION: /note="potential N-glycosylation site"  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: 372  
 OTHER INFORMATION: /note="potential N-glycosylation site"  
 US-08-868-435-31

Query Match 5.28; Score 115; DB 3; Length 466;  
 Best Local Similarity 21.1%; Pred. No. 0.0021;  
 Matches 87; Conservative 53; Mismatches 154; Indels 118; Gaps 18;

Qy 31 PSGYTLERWVILSRHVSPTKQ-----TQIMNDVTPD-----KWQWPVKAGY 74  
 Db 65 PONCKITFVQLLSHGARYPTSSKTELYSLISRIQKTATAYKGYAFKDYRYQLGAND 124  
 Qy 75 LTPGAEVLTMGFGYGDYFSLGLLAAGCPAEGGVVAADIDQRTLRITQAFLDGVAPG 134  
 Db 125 LTPFGENQMLQGIKFTNYKSLARNV-----PFVRCGSDRVIASGRLEIEGFQSA 177  
 Qy 135 CGLTVHVNQADLKTDPLFHPV-----ETGVCKLDNAQTDKAIEERLGGPLDTVS 183  
 Db 178 KVLDPH--SDKHADPTINVIIEGSPSYNNTLDGSPV-----FEDSSG--HDAQ 225  
 Qy 184 QRYAKPFAQM-----GDVLFNFAAPYCKSLQQGKTCDFAHFAANEVNVNKEGTYK 233  
 Db 226 EKPAKQAPALEIKIDHLPQVDLAVSDVPEYMDL-----CPFETLARHNTD----- 272  
 Qy 234 VTLSGPLALSSTLGEIFLLQNAQMPVAVQRLKGAENWVSLLSLH--NAQFNLAK--- 288

Db 273 -TLPSPFCALST-----QBEWQ-----AYDYQSLGKYYGNGGNGPLGPAQG 312  
 Qy 289 ---TPYIARHKGTPLLQOQIDTALTQLDAGQKLPISAQNRVLFGLGHDTNIAIAGML 344  
 Db 313 VGFVNELIARMTSPV--QDYTVNHTLDSNPATFPPLNATLYADF--SHDNTWTSIFAAL 368  
 Qy 345 G-----ADWQLPEQPD-----NTPPGGLVPELWQNPONHORYVAV 380  
 Db 369 GLYNGTAKLSTTEIKSIEETDGYSAAWTVPFGRAYIEMMQCDDSDPEPVRV 420

RESULT 13  
 US-08-744-231-31  
 ; Sequence 31, Application US/08744231  
 ; Patent No. 6358722  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Van Loon, Adolphus  
 ; APPLICANT: Mitchell, David  
 ; TITLE OF INVENTION: POLYPEPTIDES WITH PHYTASE ACTIVITY  
 ; NUMBER OF SEQUENCES: 35  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Hoffmann-La Roche Inc.  
 ; STREET: 340 Kingsland Street  
 ; CITY: Nutley  
 ; STATE: New Jersey  
 ; COUNTRY: United States of America  
 ; ZIP: 07110  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/744,231  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/424,757  
 ; FILING DATE: 18-APR-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Kass, Alan P  
 ; REGISTRATION NUMBER: 32142  
 ; REFERENCE/DOCKET NUMBER: Case Docket 9339  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (201) 235-2363  
 ; TELEFAX: (201) 235-2363  
 ; INFORMATION FOR SEQ ID NO: 31:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 466 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: 204  
 ; OTHER INFORMATION: /note="potential N-glycosylation site"  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: 269  
 ; OTHER INFORMATION: /note="potential N-glycosylation site"  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: 335  
 ; OTHER INFORMATION: /note="potential N-glycosylation site"  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: 348  
 ; OTHER INFORMATION: /note="potential N-glycosylation site"  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: 372  
 ; OTHER INFORMATION: /note="potential N-glycosylation site"

US-08-744-231-31

Query Match 5.2%; Score 115; DB 4; Length 466;  
Best Local Similarity 21.1%; Pred. No. 0.0021;  
Matches 87; Conservative 53; Mismatches 154; Indels 118; Gaps 18;

QY 31 PSGYTLERVLRSRGVRSPTKQ-----TOLMNDVTPD-----KWPQWPKVAGY 74  
DB 65 PONCKITFVQLLSRHGARYPTSSKTELXSLISRIQKTATAYKGYAFELKDYRYQLGAND 124  
QY 75 LTPRGAELVTLMGFGYGYFSLGLLAAGCPAEGGVYAQADIDQRTLTGQAFLDGVAPG 134  
DB 125 LTPFGENQMIQGIKFYKHYKSLARNV-----PFVRCSGSDRVIASGRFLIEGFQSA 177  
QY 135 CGLTVNQADLKKTDPLFHPV-----ETGVCKLDNAQTDKAIERLGGPLDVTVS 183  
DB 178 KVLDPH--SDXHDAPPTINVIIIEGSPSYNTLDTGSCPV-----FEDSSGG--HDAQ 225  
QY 184 QRYAKPEAQM-----GDVLFNFAASPYCKSLQOQKTCDFAHFAANEVNVNKEGTK 233  
DB 226 EKFAKQFAPALEKTKHLPVGLAVSDVPYMLDL-----CPFETLARNHTD----- 272  
QY 234 VTLGSLFALSSTLGEIFLLQNAQAMPEVAWQRLKGAENWVLSLSH--NAQFNLMAK--- 288  
DB 273 -TLSPPFCALST-----QBEWQ-----AYDYQSLGKYNGGNGPLGPAQ 312  
QY 289 ----TPYIARHKGTPLLQIDTALTQLDAQOQKLPISAQNRVFLFGHDTNIANIAGML 344  
DB 313 VGFVNELIARWTHSPV--QDYTVNHTLDSNPATFPLNATLYADF--SHDNTMTSIFAAL 368  
QY 345 G-----ADWQLEPEQD-----NTPPGGLVFLWQNPNDHQRVAV 380  
DB 369 GLYNGTAKLSTTEIKSIEETDGYSAAWTVFGGRAYIEMWQCDSDSEPVVRV 420

RESULT 14

US-09-273-871A-13  
; Sequence 13, Application US/09273871A  
; Patent No. 6514495  
; GENERAL INFORMATION:  
; APPLICANT: Svendsen, Allan  
; TITLE OF INVENTION: Phytase Variants  
; FILE REFERENCE: 5618.500-US  
; CURRENT APPLICATION NUMBER: US/09/273,871A  
; CURRENT FILING DATE: 1999-03-22  
; PRIOR APPLICATION NUMBER: PA 1998 00407  
; PRIOR FILING DATE: 1998-03-23  
; PRIOR APPLICATION NUMBER: PA 1998 00806  
; PRIOR FILING DATE: 1998-06-19  
; PRIOR APPLICATION NUMBER: PA 1998 01176  
; PRIOR FILING DATE: 1998-09-18  
; PRIOR APPLICATION NUMBER: PA 1999 00091  
; PRIOR FILING DATE: 1999-01-22  
; PRIOR APPLICATION NUMBER: 60/080,129  
; PRIOR FILING DATE: 1998-03-31  
; PRIOR APPLICATION NUMBER: 60/090,675  
; PRIOR FILING DATE: 1998-06-25  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13  
; LENGTH: 466  
; TYPE: PRY  
; ORGANISM: Talaromyces thermophilus  
US-09-273-871A-13

Query Match 5.2%; Score 115; DB 4; Length 466;  
Best Local Similarity 21.1%; Pred. No. 0.0021;  
Matches 87; Conservative 53; Mismatches 154; Indels 118; Gaps 18;

QY 31 PSGYTLERVLRSRGVRSPTKQ-----TOLMNDVTPD-----KWPQWPKVAGY 74  
DB 65 PONCKITFVQLLSRHGARYPTSSKTELXSLISRIQKTATAYKGYAFELKDYRYQLGAND 124

QY 75 LTPRGAELVTLMGFGYGYFSLGLLAAGCPAEGGVYAQADIDQRTLTGQAFLDGVAPG 134  
DB 125 LTPFGENQMIQGIKFYKHYKSLARNV-----PFVRCSGSDRVIASGRFLIEGFQSA 177  
QY 135 CGLTVNQADLKKTDPLFHPV-----ETGVCKLDNAQTDKAIERLGGPLDVTVS 183  
DB 178 KVLDPH--SDXHDAPPTINVIIIEGSPSYNTLDTGSCPV-----FEDSSGG--HDAQ 225  
QY 184 QRYAKPEAQM-----GDVLFNFAASPYCKSLQOQKTCDFAHFAANEVNVNKEGTK 233  
DB 226 EKFAKQFAPALEKTKHLPVGLAVSDVPYMLDL-----CPFETLARNHTD----- 272  
QY 234 VTLGSLFALSSTLGEIFLLQNAQAMPEVAWQRLKGAENWVLSLSH--NAQFNLMAK--- 288  
DB 273 -TLSPPFCALST-----QBEWQ-----AYDYQSLGKYNGGNGPLGPAQ 312  
QY 289 ----TPYIARHKGTPLLQIDTALTQLDAQOQKLPISAQNRVFLFGHDTNIANIAGML 344  
DB 313 VGFVNELIARWTHSPV--QDYTVNHTLDSNPATFPLNATLYADF--SHDNTMTSIFAAL 368  
QY 345 G-----ADWQLEPEQD-----NTPPGGLVFLWQNPNDHQRVAV 380  
DB 369 GLYNGTAKLSTTEIKSIEETDGYSAAWTVFGGRAYIEMWQCDSDSEPVVRV 420

RESULT 15

US-09-252-991A-21854  
; Sequence 21854, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 21854  
; LENGTH: 2291  
; TYPE: PRY  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-21854

Query Match 4.5%; Score 99; DB 4; Length 2291;  
Best Local Similarity 22.7%; Pred. No. 1.7;  
Matches 74; Conservative 38; Mismatches 116; Indels 98; Gaps 16;

QY 5 ENRVRLSGLVL-----MLSLAAITAPVA-----AEPGYTLERVV--ILSRHGVR 49  
DB 1992 ENERLNGCVAVGNCACAEIRAEIDAGSTALNELVARQETANPGSDSDIAVGFLMGRNVVD 2051  
QY 50 PTKQTOLMNDVTPDKV-----POMPVKAG--YLTPRGALVTLMGFGYGYDIFRSILGLAAGC 104  
DB 2052 WTTAGQLHLEQTANLWNGNPGQKEVGAYLDQTG-----FNPFGI---GV 2094  
QY 105 PAEGGVYAQADIDQRTLTGQAFLDG-----VAPGCG--LTVHQAQDLKTDPLF 152  
DB 2095 PAMGGAAG-----KVTAKALMNALKAGEIPKGEVAFGKANLPFGALADAEAGMPT 2146  
QY 153 HPVE-----TGVCCKLDNAQTDKAIERLGGPLDVTVSQRYAKPFAQMGVILNFAASPY 204  
DB 2147 HPVKLAAKATGTAGKIKIEAGAIIPANEVRAGQGL-----SGLGYDVTHQTAS 2195  
QY 205 CKSLQOQKTCDFAHFAANEVNV-----NKEGKVT-----LSGPLA-----LS 243  
DB 2196 AKGIQSQ--RTADLHVGLGSLDIVTTPKNDLPNTKIVRAIEKKSNQAGVLVQADLPSTMS 2254  
QY 244 STLGEIFLLQNAQAMPEVAWQRLKGA 269

Db 2255 SIAARMWGKINAOSIKITIPFOKPDGS 2280

Search completed: April 30, 2004, 12:43:16  
Job time : 18.8293 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 12:41:31 ; Search time 40.9756 Seconds  
(without alignments)  
2841.193 Million cell updates/sec

Title: US-10-021-723B-4  
Perfect score: 2200  
Sequence: 1 MSVLENRVRLSGVLMLSGLAITAPVAAPPSGYTLERVILSRHGVRSPKOTQLMNDV 60

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1138130 seqs, 27719581 residues

Total number of hits satisfying chosen parameters: 1138120

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Published Applications AA:  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US05\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2200	100.0	420	14	US-10-021-723A-4
2	2193	99.7	441	14	US-10-021-723A-12
3	2183	99.2	441	12	US-10-282-122A-77792
4	2175	98.9	441	14	US-10-021-723A-2
5	948.5	43.1	436	14	US-10-156-660-2
6	943	42.9	410	14	US-10-021-723A-13
7	943	42.9	412	14	US-10-021-723A-15
8	943	42.9	412	14	US-10-334-672-1
9	943	42.9	412	14	US-10-334-671-1
10	939.5	42.7	432	9	US-09-866-379-8
11	939.5	42.7	432	12	US-10-282-122A-43351
12	939.5	42.7	432	14	US-10-156-660-4
13	932.5	42.4	440	9	US-09-777-566A-2
14	932.5	42.4	440	9	US-09-866-379-2
15	932.5	42.4	440	14	US-10-034-985-2

16	932.5	42.4	440	15	US-10-430-356-2	Sequence 2, Appli
17	928.5	42.2	432	15	US-10-284-962-3	Sequence 3, Appli
18	928.5	42.2	432	15	US-10-284-962-14	Sequence 14, Appli
19	928.5	42.2	433	14	US-10-266-041-1	Sequence 1, Appli
20	928.5	42.2	476	14	US-10-021-723A-10	Sequence 10, Appli
21	921.5	41.9	430	9	US-09-866-379-10	Sequence 10, Appli
22	920.5	41.8	432	15	US-10-284-962-5	Sequence 5, Appli
23	886.5	40.3	432	14	US-10-021-723A-16	Sequence 16, Appli
24	567	25.8	318	14	US-10-021-723A-14	Sequence 14, Appli
25	504	22.9	409	14	US-10-021-723A-8	Sequence 8, Appli
26	482	21.9	421	14	US-10-021-723A-6	Sequence 6, Appli
27	126	5.7	99	12	US-10-282-122A-59099	Sequence 59099, A
28	115	5.2	443	12	US-10-442-538-112	Sequence 112, App
29	115	5.2	443	12	US-10-442-538-134	Sequence 134, App
30	115	5.2	443	12	US-10-442-538-157	Sequence 157, App
31	115	5.2	450	14	US-10-062-848-9	Sequence 9, Appli
32	115	5.2	466	13	US-10-083-452-13	Sequence 13, Appli
33	113	5.1	1169	12	US-10-282-122A-51464	Sequence 51464, A
34	108	4.9	26926	9	US-09-759-508B-2	Sequence 2, Appli
35	101.5	4.6	479	12	US-10-282-122A-54804	Sequence 54804, A
36	99.5	4.5	413	15	US-10-369-493-3577	Sequence 3577, Ap
37	98.5	4.5	1224	12	US-10-282-122A-68248	Sequence 68248, A
38	98	4.5	356	12	US-10-424-599-203839	Sequence 203839,
39	98	4.5	436	15	US-10-369-493-6811	Sequence 6811, Ap
40	97.5	4.4	348	12	US-10-282-122A-60983	Sequence 60983, A
41	97.5	4.4	466	15	US-10-369-493-10596	Sequence 10596, A
42	97.5	4.4	4928	16	US-10-329-148A-5	Sequence 5, Appli
43	97	4.4	338	14	US-10-156-761-12751	Sequence 12751, A
44	96.5	4.4	327	15	US-10-369-493-6327	Sequence 6327, Ap
45	96	4.4	445	14	US-10-156-761-13900	Sequence 13900, A

## ALIGNMENTS

RESULT 1  
US-10-021-723A-4  
; Sequence 4, Application US/10021723A  
; Publication No. US20030101476A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay  
; APPLICANT: Mathur, Eric  
; APPLICANT: Richardson, Toby  
; APPLICANT: Robertson, Dan  
; APPLICANT: Barton, Nelson  
; TITLE OF INVENTION: Recombinant Phytases and Uses Thereof  
; FILE REFERENCE: 112766.140 (DIV-016CIP)  
; CURRENT APPLICATION NUMBER: US/10/021,723A  
; CURRENT FILING DATE: 2002-10-23  
; PRIOR APPLICATION NUMBER: US 60/255,090  
; PRIOR FILING DATE: 2000-12-12  
; NUMBER OF SEQ ID NOS: 74  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 420  
; TYPE: PRT  
; ORGANISM: Versinia pestis  
US-10-021-723A-4

Query Match	100.0%;	Score 2200;	DB 14;	Length 420;
Best Local Similarity	100.0%;	Pred. No. 4.1e-207;		
Matches 420;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MSVLENRVRLSGVLMLSGLAITAPVAAPPSGYTLERVILSRHGVRSPKOTQLMNDV	60	
Db	1	MSVLENRVRLSGVLMLSGLAITAPVAAPPSGYTLERVILSRHGVRSPKOTQLMNDV	60	
Qy	61	TPDKWPQWPVKAGYLPTRGAEVLTLMGGFYGDYFRSLGLLAAGCPAEGGYAQAIDQRT	120	
Db	61	TPDKWPQWPVKAGYLPTRGAEVLTLMGGFYGDYFRSLGLLAAGCPAEGGYAQAIDQRT	120	
Qy	121	RLTGQAFLDGVAFCGLTVNQADLKTDLFPVETGVCKLNAQTDKAIEERLGGPLD	180	

Db 121 RLTAQAFDVGAPCGGLTVHQAADLKKTDPLFHPVETGVCKLDNAQTDKAIEERLGGPLD 180  
Qy 181 TVSORYAKPAQMGDVNLNFAASPYCKSLQOQGTCTDFAHFAANEVNVNKEGTVTLGGPL 240  
Db 181 TVSORYAKPAQMGDVNLNFAASPYCKSLQOQGTCTDFAHFAANEVNVNKEGTVTLGGPL 240  
Qy 241 ALSSTLGEIIFLLQNAQAMPEVAVQRLKGAENWVLSLHNAQFNLMNAKTPYIARHKGTP 300  
Db 241 ALSSTLGEIIFLLQNAQAMPEVAVQRLKGAENWVLSLHNAQFNLMNAKTPYIARHKGTP 300  
Qy 301 LQOITDALTQLDQAQOKLPIAQNRYVFLGGHDTNIAIAGMLGADWQLPEQPDNTPPG 360  
Db 301 LQOITDALTQLDQAQOKLPIAQNRYVFLGGHDTNIAIAGMLGADWQLPEQPDNTPPG 360  
Qy 361 GGLVFELWQPNDRHQRVAVKMFYQTMQDLRNEAKLKLKNNPAGIISVAVAGCENNGDDK 420  
Db 361 GGLVFELWQPNDRHQRVAVKMFYQTMQDLRNEAKLKLKNNPAGIISVAVAGCENNGDDK 420

RESULT 2

US-10-021-723A-12  
; Sequence 12, Application US/10021723A  
; Publication No. US20030101476A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay  
; APPLICANT: Mathur, Eric  
; APPLICANT: Richardson, Toby  
; APPLICANT: Robertson, Dan  
; APPLICANT: Barton, Nelson  
; TITLE OF INVENTION: Recombinant Phytases and Uses Thereof  
; FILE REFERENCE: 112765.140 (DIV-016CIP)  
; CURRENT APPLICATION NUMBER: US/10/021,723A  
; CURRENT FILING DATE: 2002-10-23  
; PRIOR FILING DATE: 2000-12-12  
; PRIOR FILING DATE: 2000-12-12  
; NUMBER OF SEQ ID NOS: 74  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 441  
; TYPE: PRT  
; ORGANISM: Yersinia pestis  
US-10-021-723A-12

Query Match 99.7%; Score 2193; DB 14; Length 441;  
Best Local Similarity 99.8%; Pred. No. 2.1e-206;  
Matches 419; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MSVLENRVLSGLVLMISGLAAITAPVAAPSGYTLERVILSRHGVRSPTKOTQLMNDV 60  
Db 1 MSGLENRVLSGLVLMISGLAAITAPVAAPSGYTLERVILSRHGVRSPTKOTQLMNDV 60  
Qy 61 TPDKWPQWPVKAGYLTFRGAEVLTLMGFGYDYSRSLGGLAAGCPAEGGVYAADIDQRT 120  
Db 61 TPDKWPQWPVKAGYLTFRGAEVLTLMGFGYDYSRSLGGLAAGCPAEGGVYAADIDQRT 120  
Qy 121 RLTAQAFDVGAPCGGLTVHQAADLKKTDPLFHPVETGVCKLDNAQTDKAIEERLGGPLD 180  
Db 121 RLTAQAFDVGAPCGGLTVHQAADLKKTDPLFHPVETGVCKLDNAQTDKAIEERLGGPLD 180  
Qy 181 TVSORYAKPAQMGDVNLNFAASPYCKSLQOQGTCTDFAHFAANEVNVNKEGTVTLGGPL 240  
Db 181 TVSORYAKPAQMGDVNLNFAASPYCKSLQOQGTCTDFAHFAANEVNVNKEGTVTLGGPL 240  
Qy 241 ALSSTLGEIIFLLQNAQAMPEVAVQRLKGAENWVLSLHNAQFNLMNAKTPYIARHKGTP 300  
Db 241 ALSSTLGEIIFLLQNAQAMPEVAVQRLKGAENWVLSLHNAQFNLMNAKTPYIARHKGTP 300  
Qy 301 LQOITDALTQLDQAQOKLPIAQNRYVFLGGHDTNIAIAGMLGADWQLPEQPDNTPPG 360  
Db 301 LQOITDALTQLDQAQOKLPIAQNRYVFLGGHDTNIAIAGMLGADWQLPEQPDNTPPG 360  
Qy 361 GGLVFELWQPNDRHQRVAVKMFYQTMQDLRNEAKLKLKNNPAGIISVAVAGCENNGDDK 420

Db 361 GGLVFELWQPNDRHQRVAVKMFYQTMQDLRNEAKLKLKNNPAGIISVAVAGCENNGDDK 420  
RESULT 3  
US-10-282-122A-77792  
; Sequence 77792, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsuo  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 77792  
; LENGTH: 441  
; TYPE: PRT  
; ORGANISM: Yersinia pestis  
US-10-282-122A-77792

Query Match 99.2%; Score 2183; DB 12; Length 441;  
Best Local Similarity 99.3%; Pred. No. 2.1e-205;  
Matches 417; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MSVLENRVLSGLVLMISGLAAITAPVAAPSGYTLERVILSRHGVRSPTKOTQLMNDV 60  
Db 1 MSVLENRVLSGLVLMISGLAAITAPVAAPSGYTLERVILSRHGVRSPTKOTQLMNDV 60  
Qy 61 TPDKWPQWPVKAGYLTFRGAEVLTLMGFGYDYSRSLGGLAAGCPAEGGVYAADIDQRT 120  
Db 61 TPDKWPQWPVKAGYLTFRGAEVLTLMGFGYDYSRSLGGLAAGCPAEGGVYAADIDQRT 120  
Qy 121 RLTAQAFDVGAPCGGLTVHQAADLKKTDPLFHPVETGVCKLDNAQTDKAIEERLGGPLD 180  
Db 121 RLTAQAFDVGAPCGGLTVHQAADLKKTDPLFHPVETGVCKLDNAQTDKAIEERLGGPLD 180  
Qy 181 TVSORYAKPAQMGDVNLNFAASPYCKSLQOQGTCTDFAHFAANEVNVNKEGTVTLGGPL 240  
Db 181 TVSORYAKPAQMGDVNLNFAASPYCKSLQOQGTCTDFAHFAANEVNVNKEGTVTLGGPL 240



QY 241 ALSSTLGEIFLLQNAQAMPEVAWQRLKGAENWVLSLSLHNAQFNLMAKTPYIAHKGTPPL 300  
Db 241 ALSSTLGEIFLLQNAQAMPEVAWQRLKGAENWVLSLSLHNAQFNLMAKTPYIAHKGTPPL 300  
QY 301 LQOITDALTQLDAGQKLPISAQNRVFLGSHDNTNIAIAGMLGADWQLEQPDNTPPG 360  
Db 301 LQOITDALTQLDAGQKLPISAQNRVFLGSHDNTNIAIAGMLGADWQLEQPDNTPPG 360  
QY 361 GGLVFLWQNPNDHQRVYAVKMFYQTMQDLRNAEKLKLNPNAGIISVAVAGCENNGDDK 420  
Db 361 GGLVFLWQNPNDHQRVYAVKMFYQTMQDLRNAEKLKLNPNAGIISVAVAGCENNGDDK 420

RESULT 4

US-10-021-723A-2  
; Sequence 2, Application US/10021723A  
; Publication No. US20030101476A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay  
; APPLICANT: Mathur, Eric  
; APPLICANT: Richardson, Toby  
; APPLICANT: Robertson, Dan  
; APPLICANT: Barton, Nelson  
; TITLE OF INVENTION: Recombinant Phycases and Uses Thereof  
; FILE REFERENCE: 112766.140 (DIV-016CIP)  
; CURRENT APPLICATION NUMBER: US/10/021,723A  
; CURRENT FILING DATE: 2002-10-23  
; PRIOR APPLICATION NUMBER: US 60/255,090  
; PRIOR FILING DATE: 2000-12-12  
; NUMBER OF SEQ ID NOS: 74  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 441  
; TYPE: PRT  
; ORGANISM: Yersinia pestis  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: 157, 163, 164, 174  
; OTHER INFORMATION: Xaa = Any Amino Acid  
US-10-021-723A-2

Query Match 98.9%; Score 2175; DB 14; Length 441;  
Best Local Similarity 99.0%; Pred. No. 1.3e-204;  
Matches 416; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSVLENRVLSGLVLMGLAAITAPVAAPSGYTLERVILSRHGVSPKQTLQMDV 60  
Db 1 MSVLENRVLSGLVLMGLAAITAPVAAPSGYTLERVILSRHGVSPKQTLQMDV 60  
QY 61 TPKKPPQVKAAGLTPRGAELVTLGGFYGDYFRSLGLLAAGCPAEGGVYAADIDORT 120  
Db 61 TPKKPPQVKAAGLTPRGAELVTLGGFYGDYFRSLGLLAAGCPAEGGVYAADIDORT 120  
QY 121 RLTGQAFLDGVAAGCGLTVHQAADLKTDPFLFHPVETGVCKLDAQTDKAIEERLGGPLD 180  
Db 121 RLTGQAFLDGVAAGCGLTVHQAADLKTDPFLFHPVEXGVCKLXXAQTDKAIEERLGGPLD 180  
QY 181 TVSQRYAKPFAQMGDVLFNFAASPYCKSLQOQKTCDFAHFAANEVNVNKEGTVTLGGPL 240  
Db 181 TVSQRYAKPFAQMGDVLFNFAASPYCKSLQOQKTCDFAHFAANEVNVNKEGTVTLGGPL 240  
QY 241 ALSSTLGEIFLLQNAQAMPEVAWQRLKGAENWVLSLSLHNAQFNLMAKTPYIAHKGTPPL 300  
Db 241 ALSSTLGEIFLLQNAQAMPEVAWQRLKGAENWVLSLSLHNAQFNLMAKTPYIAHKGTPPL 300  
QY 301 LQOITDALTQLDAGQKLPISAQNRVFLGSHDNTNIAIAGMLGADWQLEQPDNTPPG 360  
Db 301 LQOITDALTQLDAGQKLPISAQNRVFLGSHDNTNIAIAGMLGADWQLEQPDNTPPG 360  
QY 361 GGLVFLWQNPNDHQRVYAVKMFYQTMQDLRNAEKLKLNPNAGIISVAVAGCENNGDDK 420  
Db 361 GGLVFLWQNPNDHQRVYAVKMFYQTMQDLRNAEKLKLNPNAGIISVAVAGCENNGDDK 420

RESULT 5

US-10-156-660-2  
; Sequence 2, Application US/10156660  
; Publication No. US20030101958A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay M.  
; APPLICANT: Kretz, Keith  
; APPLICANT: Gray, Kevin A.  
; APPLICANT: Barton, Nelson R.  
; APPLICANT: Garrett, James B.  
; APPLICANT: O'Donoghue, Eileen  
; APPLICANT: Mathur, Eric J.  
; TITLE OF INVENTION: PHYTASES, NUCLEIC ACIDS ENCODING THEM  
; TITLE OF INVENTION: AND METHODS FOR MAKING AND USING THEM  
; FILE REFERENCE: 09010-029007  
; CURRENT APPLICATION NUMBER: US/10/156,660  
; CURRENT FILING DATE: 2002-10-01  
; PRIOR APPLICATION NUMBER: US 09/866,379  
; PRIOR FILING DATE: 2001-05-24  
; PRIOR APPLICATION NUMBER: US 09/580,515  
; PRIOR FILING DATE: 2000-05-25  
; PRIOR APPLICATION NUMBER: US 09/318,528  
; PRIOR FILING DATE: 1999-05-25  
; PRIOR APPLICATION NUMBER: US 09/291,931  
; PRIOR FILING DATE: 1999-04-13  
; PRIOR APPLICATION NUMBER: US 09/259,214  
; PRIOR FILING DATE: 1999-03-01  
; PRIOR APPLICATION NUMBER: US 08/910,798  
; PRIOR FILING DATE: 1997-08-13  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 436  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: modified phytase enzyme  
US-10-156-660-2

Query Match 43.1%; Score 948.5; DB 14; Length 436;  
Best Local Similarity 48.2%; Pred. No. 4e-84;  
Matches 197; Conservative 62; Mismatches 135; Indels 15; Gaps 7;

QY 13 LVLMGLAAITAPVA---AEPGSGYTLERVILSRHGVSPKQTLQMDVTPDKWPQP 69  
Db 5 LIPFLSLLIPLTPQSAFAQSEPE-LKLESVIVSRHGVRAPTKATQMDVTPDAWPTWP 63  
QY 70 VKAGYLTTPRGAELVTLGGFYGDYFRSLGLLAAGCPAEGGVYAADIDORTLTGQAF 128  
Db 64 VKUGELTPRGGELIATLGHYWRQLVADGLLPKCGCPQSGQVAITADVERTKTGEAFA 123  
QY 129 DGVAAGCGLTVHQAADLKTDPFLFHPVETGVCKLDAQTDKAIEERLGGPLDTPVQRYAK 188  
Db 124 AGLAPDCAITVHTQADTSFPDPLFNLKTVGQCLDQNVNVDAILERAGGSIADFTGHYQT 183  
QY 189 PFAQMGDVLFNFAASPYCKSLQOQKTCDFAHFAANEVNVNKEGTVTLGGPLSLSTLGE 248  
Db 184 AFRELERVILFPOSNCLKREKQDESCSLTQALPSLKVSAAD--CVSLTGAVSLASMLTPE 241  
QY 249 IFLLQNAQAMPEVAWQRLKGAENWVLSLSLHNAQFNLMAKTPYIAHKGTPPLIQQIDTAL 308  
Db 242 IFLLQNAQAMPEVPGWGRITDSHQWNVLSLSHNAQFLLQRTPEVARSRAATPLDLIKTAL 301  
QY 309 T----LQIDAGQKLPISAQNRVFLGSHDNTNIAIAGMLGADWQLEQPDNTPPGGLVFP 365  
Db 302 TPHPKQKQYGVTLPTS----VLFIAGHDNTLANLGGALELANWTLPGQPDNTPPGGELVFP 357  
QY 366 ELWQNPNDHQRVYAVKMFYQTMQDLRNAEKLKLNPNAGIISVAVAGCE 414  
Db 358 ERWRRLSDNSQWISVLFQTLQQRDKTFLSL-NTFPGEVKLTLAGCE 405

```
RESULT 6
US-10-021-723A-13
; Sequence 13, Application US/10021723A
; Publication No. US20030101476A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Mathur, Eric
; APPLICANT: Richardson, Toby
; APPLICANT: Robertson, Dan
; APPLICANT: Barton, Nelson
; TITLE OF INVENTION: Recombinant Phytases and Uses Thereof
; FILE REFERENCE: 112766.140 (DIV-016CIP)
; CURRENT APPLICATION NUMBER: US/10/021,723A
; CURRENT FILING DATE: 2002-10-23
; PRIOR APPLICATION NUMBER: US 60/255,090
; PRIOR FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 410
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: amino acid sequence of a phytase enzyme with
; OTHER INFORMATION: predicted glycosylation sites
US-10-021-723A-13

Query Match 42.9%; Score 943; DB 14; Length 410;
Best Local Similarity 49.0%; Pred. No. 1.3e-83;
Matches 191; Conservative 60; Mismatches 127; Indels 12; Gaps 6;

QY 29 AEPGTYTLERVILSRHGVSPKQTLQMDVTPDKWPQMPVKGAYLTTPRGAEIVLMGG 88
Db 2 SEPE-LKLESVIVSRHGVRAFTKATQMQDVTDPDAMPTWFKLGELTPRGELIAYLGH 60
QY 89 FYGDYFRLSGLL-AAGCPAGGVYAQADIDORTLTQAFLDGVAPCGLTVHNOADLKK 147
Db 61 YWRQLVADGLLPKCGCPQSGQVAIIADVDERTRKTGEAFAGLAPDCAITVHTQADTSS 120
QY 148 TDPLFHPVETGVCKLDNAQTDKAEERLGGPLDTSQRYAKPFAQMGDVILNFAASPYCKS 207
Db 121 PDPLFNLKTVGCQLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLNFPQSNCLK 180
QY 208 LQOQKTCDFAHFAANEVNVKGTKVTLSGPLALSTLGEIIFLQNAQAMPEVAVQRLK 267
Db 181 REKQDESCSLTQALPSELKVSAD--CVSLTGAVSLASMLTEIFLQQAQMPPEPGWGRIT 238
QY 268 GAENWVLSLSHNAQFNLMAKTPYIARHKGTPLLQQIDTALT---LQIDAQGGKLPISAQ 324
Db 239 DSHQWNTLLSHNAQFLLQRTPEVARSRAATPLDLIKTALTTPHPKQAYGVTLPST 296
QY 325 NRVLFLGHDNTNANIAGMLGADWOLPEQDNTPPGGGLVFELWQNDPNHQRVAVVMPY 384
Db 297 --VLFIAGHDNTLANLGALXNLWTLPGQDNTPPGGELVFERWRRLSDNSQWIQVSLVF 354
QY 385 QTMQLRNAEKLDKNNPAGIISVAVAGCE 414
Db 355 QTLQMRDKTPLSL-NTPPGEVKTLAGCE 383

RESULT 8
US-10-334-672-1
; Sequence 1, Application US/10334672
; Publication No. US20030157646A1
; GENERAL INFORMATION:
; APPLICANT: Lananan, Mike
; APPLICANT: Koepf, Edward
; APPLICANT: Kretz, Keith
; TITLE OF INVENTION: Microbially-Expressed Thermotolerant Phytase For Animal Feed
; FILE REFERENCE: SYNG-P01-001
; CURRENT APPLICATION NUMBER: US/10/334,672
; CURRENT FILING DATE: 2002-12-30
; PRIOR APPLICATION NUMBER: 60/344,523
; PRIOR FILING DATE: 2001-12-28
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: No. US20030157646A19X Phytase
US-10-334-672-1

Query Match 42.9%; Score 943; DB 14; Length 412;
Best Local Similarity 49.0%; Pred. No. 1.3e-83;
```



```
QY 309 T---LQDAQOGKLPISANRVLFLGGHDTNANIAGMLGADWOLPEQDNTTPGGGLVF 365
Db 302 TPHPFKQKQXGVTLPTS-----VLFIAHDTNLANLGALLENWTLPGQDNTTPGGGLVF 357
QY 366 ELWQNDPNHORYVAVKMFQVMDQURNAEKLDKNNPAGIISVAVAGE 414
Db 358 ERWRRLSDNSQWISQVSLVFTLQQRMDKTPLSL-NTPPGEVKLTLAGCE 405

RESULT 11
US-10-282-122A-43351
; Sequence 4351, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Haselbeck, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELIWA 0343
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 43351
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-282-122A-43351

Query Match 42.7%; Score 939.5; DB 12; Length 432;
Best Local Similarity 47.9%; Pred. No. 3e-83;
Matches 196; Conservative 61; Mismatches 137; Indels 15; Gaps 7;

QY 13 LVLMGLAATAAPVA---AEPSCYTLERWVLSRHGVRSPTKQTQLMNDVTPDKWPWP 69
Db 5 LIPFLSLIPLTPQSAPQSEPE-LKLESVWVSRHGVRAPTKATQMLQDVTTPDWTWP 63
QY 70 VKAGYLTFRGAEVLTMGGFYGYDFRSLGLLA-AGCPAEGGVYAQADIDQRTLTGOAFL 128
Db 64 VXLGWLTPRGELIAYLGHVQORLVADGLLAKKGCPSQGVATADVDERTKTEGAPA 123
QY 129 DGVAFCGLTVHQAOLKKTDPLEPHVETGVCKLDNAOTDKAIEERLGGPLDVTVSQYAK 188
Db 124 AGLAPDCALTHTQADTSSPDLPFNPLKTVGCQOLDNANVTDAILSRAGGSIAIDPTGHRQT 183
QY 189 PFAQMGDVILNFAASPYCKSLQOQKTCDFAHFAANEVNVNKGKTKVLSGFLSSTLGE 248
```

```
Db 124 AGLAPDCALTHTQADTSSPDLPFNPLKTVGCQOLDNANVTDAILSRAGGSIAIDPTGHRQT 183
QY 189 PFAQMGDVILNFAASPYCKSLQOQKTCDFAHFAANEVNVNKGKTKVLSGFLSSTLGE 248
Db 184 AFRELVLNFPQSNLCKREKQDESCSLTQALPSLKVGSAD--NVSLTGAIVSLASMLTE 241
QY 249 IFLLQNAQAMPEVAMORUKGAENVVLSLHNAQFNIAWAKTPYIARHKGTPLLQQIDTAL 308
Db 242 IFLLQQAQGMPEPGWGRITDSHQNWTLLSLHNAQFYLLQRTPEVARSBATFLDLIKTAL 301
QY 309 T---LQDAQOGKLPISANRVLFLGGHDTNANIAGMLGADWOLPEQDNTTPGGGLVF 365
Db 302 TPHPFKQKQXGVTLPTS-----VLFIAHDTNLANLGALLENWTLPGQDNTTPGGGLVF 357
QY 366 ELWQNDPNHORYVAVKMFQVMDQURNAEKLDKNNPAGIISVAVAGE 414
Db 358 ERWRRLSDNSQWISQVSLVFTLQQRMDKTPLSL-NTPPGEVKLTLAGCE 405

RESULT 12
US-10-156-660-4
; Sequence 4, Application US/10156660
; Publication No. US20030103958A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay M.
; APPLICANT: Kretz, Keith
; APPLICANT: Gray, Kevin A.
; APPLICANT: Barton, Nelson R.
; APPLICANT: Garrett, James B.
; APPLICANT: O'Donoghue, Eileen
; APPLICANT: Mathur, Eric J.
; TITLE OF INVENTION: PHYTASES, NUCLEIC ACIDS ENCODING THEM
; TITLE OF INVENTION: AND METHODS FOR MAKING AND USING THEM
; FILE REFERENCE: 09010-029007
; CURRENT APPLICATION NUMBER: US/10/156,660
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: US 09/866,379
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 09/580,515
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: US 09/259,214
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: US 08/910,798
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-156-660-4

Query Match 42.7%; Score 939.5; DB 14; Length 432;
Best Local Similarity 47.9%; Pred. No. 3e-83;
Matches 196; Conservative 61; Mismatches 137; Indels 15; Gaps 7;

QY 13 LVLMGLAATAAPVA---AEPSCYTLERWVLSRHGVRSPTKQTQLMNDVTPDKWPWP 69
Db 5 LIPFLSLIPLTPQSAPQSEPE-LKLESVWVSRHGVRAPTKATQMLQDVTTPDWTWP 63
QY 70 VKAGYLTFRGAEVLTMGGFYGYDFRSLGLLA-AGCPAEGGVYAQADIDQRTLTGOAFL 128
Db 64 VXLGWLTPRGELIAYLGHVQORLVADGLLAKKGCPSQGVATADVDERTKTEGAPA 123
QY 129 DGVAFCGLTVHQAOLKKTDPLEPHVETGVCKLDNAOTDKAIEERLGGPLDVTVSQYAK 188
Db 124 AGLAPDCALTHTQADTSSPDLPFNPLKTVGCQOLDNANVTDAILSRAGGSIAIDPTGHRQT 183
QY 189 PFAQMGDVILNFAASPYCKSLQOQKTCDFAHFAANEVNVNKGKTKVLSGFLSSTLGE 248
```

Db 184 AFRERLVNFFQNSNCLKREKQDSCSITQALPSELKVSAD--NVSUTGAVSLASMLTE 241  
 Qy 249 IFLLQNAQAMPEVAVQRLKGAENWVSLSLHNAQFNLMKTPYIARHKGTPILLOQIDTAL 308  
 Db 242 IFLLQNAQAMPEVAVQRLKGAENWVSLSLHNAQFNLMKTPYIARHKGTPILLOQIDTAL 301  
 Qy 309 T---LQIDAQOQKLPISAGNRVFLFGGHDNTNIAAGMLGADWQLPQPDNTPPGGGLVF 365  
 Db 302 TPHPKQKQAYGVTLPTS-----VLFIAGHDTNLANLGGALELNWTLPGQPDNTPPGGGLVF 357  
 Qy 366 ELQNPNDHQRVYAVVMFYQTMQDLQNAEKLDLKNPAGIISVAVAGCE 414  
 Db 358 ERWRRLSDNSQWISQVSLVFQTLQOMDKTPLSL-NTPPGVKLTLAGCE 405

RESULT 13  
 US-09-777-566A-2  
 ; Sequence 2, Application US/09777566A  
 ; Patent No. US20010055788A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DIVERSA CORPORATION  
 ; APPLICANT: SHORT, Jay  
 ; APPLICANT: KREITZ, Keith  
 ; APPLICANT: BARTON, Nelson  
 ; APPLICANT: GARRETT, James  
 ; APPLICANT: O'DONOGHUE, Eileen  
 ; TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYTASES AND USES THEREOF  
 ; FILE REFERENCE: DIVER1370-6  
 ; CURRENT APPLICATION NUMBER: US/09/777,566A  
 ; PRIOR FILING DATE: 2001-06-11  
 ; PRIOR APPLICATION NUMBER: US 09/318,528  
 ; PRIOR FILING DATE: 1999-05-25  
 ; PRIOR APPLICATION NUMBER: US 09/291,931  
 ; PRIOR FILING DATE: 1999-04-13  
 ; PRIOR APPLICATION NUMBER: US 09/259,214  
 ; PRIOR FILING DATE: 1999-03-01  
 ; PRIOR APPLICATION NUMBER: US 08/910,798  
 ; PRIOR FILING DATE: 1997-08-13  
 ; NUMBER OF SEQ ID NOS: 4  
 ; SOFTWARE: Patent in version 3.0  
 ; SEQ ID NO 2  
 ; LENGTH: 440  
 ; TYPE: PRT  
 ; ORGANISM: Escherichia coli  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)..(1323)  
 ; OTHER INFORMATION: n is any nucleotide  
 US-09-777-566A-2

Query Match 42.4%; Score 932.5; DB 9; Length 440;  
 Best Local Similarity 47.7%; Pred. No. 1.5e-82;  
 Matches 195; Conservative 61; Mismatches 138; Indels 15; Gaps 7;

Qy 13 LVLMLSGLAATAPVA---AEPGTYLTVSRVILSHGHVSPKTKQTLQMDVTPDKWPWP 69  
 Db 5 LIPFLSLIPLTPQSAFAQSEPE-LKLESVTVSRHGVRAPTKATQMLQMDVTPDAMPWP 63  
 Qy 70 VKAGYLTTPRGAEVLTMGGFYGDYFRSLGLLA-ACPAEGGVYAAQDIQRTLTQGAFL 128  
 Db 64 VKLGMWTPRGGLIAYLGHYQRLVADGLLAKKGCPSQGVATIAVDERTKRTGEAFA 123  
 Qy 129 DGVAPCGGLTVHQAADLKKTDPLFHPVETGVCKLDNAQTDKATEERLGGPLDVTVSQRYAK 188  
 Db 124 AGLAPDCAITVHTQADTSSPDPLFNPLKTVGVCQLDNANVTDAISRAGGSIAFTGHRQT 183  
 Qy 189 PFAQMGDLVNFAPASPYCKSLQOQKTCDFAHFAANVAVNKEGTVKTLGSLPLASSTIGE 248  
 Db 184 APELERVLNFPQSNCLKREKQDSCSITQALPSELKVSAD--NVSUTGAVSLASMLTE 241  
 Qy 249 IFLLQNAQAMPEVAVQRLKGAENWVSLSLHNAQFNLMKTPYIARHKGTPILLOQIDTAL 308  
 Db 242 IFLLQNAQAMPEVAVQRLKGAENWVSLSLHNAQFNLMKTPYIARHKGTPILLOQIDTAL 301  
 Qy 309 T---LQIDAQOQKLPISAGNRVFLFGGHDNTNIAAGMLGADWQLPQPDNTPPGGGLVF 365  
 Db 302 TPHPKQKQAYGVTLPTS-----VLFIAGHDTNLANLGGALELNWTLPGQPDNTPPGGGLVF 357  
 Qy 366 ELQNPNDHQRVYAVVMFYQTMQDLQNAEKLDLKNPAGIISVAVAGCE 414  
 Db 358 ERWRRLSDNSQWISQVSLVFQTLQOMDKTPLSL-NTPPGVKLTLAGCE 405

Db 302 TPHPKQKQAYGVTLPTS-----VLFIAGHDTNLANLGGALELNWTLPGQPDNTPPGGGLVF 357  
 Qy 366 ELQNPNDHQRVYAVVMFYQTMQDLQNAEKLDLKNPAGIISVAVAGCE 414  
 Db 358 ERWRRLSDNSQWISQVSLVFQTLQOMDKTPLSL-NTPPGVKLTLAGCE 405

RESULT 14  
 US-09-866-379-2  
 ; Sequence 2, Application US/09866379  
 ; Patent No. US20020136754A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DIVERSA CORPORATION  
 ; APPLICANT: SHORT, Jay  
 ; APPLICANT: KREITZ, Keith  
 ; APPLICANT: BARTON, Nelson  
 ; APPLICANT: GARRETT, James  
 ; APPLICANT: O'DONOGHUE, Eileen  
 ; TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYTASES AND USES THEREOF  
 ; FILE REFERENCE: DIVER1370-7  
 ; CURRENT APPLICATION NUMBER: US/09/866,379  
 ; PRIOR FILING DATE: 2001-05-24  
 ; PRIOR APPLICATION NUMBER: US 09/580,515  
 ; PRIOR FILING DATE: 2000-05-25  
 ; PRIOR APPLICATION NUMBER: US 09/318,528  
 ; PRIOR FILING DATE: 1999-05-25  
 ; PRIOR APPLICATION NUMBER: US 09/291,931  
 ; PRIOR FILING DATE: 1999-04-13  
 ; PRIOR APPLICATION NUMBER: US 09/259,214  
 ; PRIOR FILING DATE: 1999-03-01  
 ; PRIOR APPLICATION NUMBER: US 08/910,798  
 ; PRIOR FILING DATE: 1997-08-13  
 ; NUMBER OF SEQ ID NOS: 10  
 ; SOFTWARE: Patent in version 3.1  
 ; SEQ ID NO 2  
 ; LENGTH: 440  
 ; TYPE: PRT  
 ; ORGANISM: Escherichia coli  
 ; OTHER INFORMATION: n is any nucleotide  
 US-09-866-379-2

Query Match 42.4%; Score 932.5; DB 9; Length 440;  
 Best Local Similarity 47.7%; Pred. No. 1.5e-82;  
 Matches 195; Conservative 61; Mismatches 138; Indels 15; Gaps 7;

Qy 13 LVLMLSGLAATAPVA---AEPGTYLTVSRVILSHGHVSPKTKQTLQMDVTPDKWPWP 69  
 Db 5 LIPFLSLIPLTPQSAFAQSEPE-LKLESVTVSRHGVRAPTKATQMLQMDVTPDAMPWP 63  
 Qy 70 VKAGYLTTPRGAEVLTMGGFYGDYFRSLGLLA-ACPAEGGVYAAQDIQRTLTQGAFL 128  
 Db 64 VKLGMWTPRGGLIAYLGHYQRLVADGLLAKKGCPSQGVATIAVDERTKRTGEAFA 123  
 Qy 129 DGVAPCGGLTVHQAADLKKTDPLFHPVETGVCKLDNAQTDKATEERLGGPLDVTVSQRYAK 188  
 Db 124 AGLAPDCAITVHTQADTSSPDPLFNPLKTVGVCQLDNANVTDAISRAGGSIAFTGHRQT 183  
 Qy 189 PFAQMGDLVNFAPASPYCKSLQOQKTCDFAHFAANVAVNKEGTVKTLGSLPLASSTIGE 248  
 Db 184 APELERVLNFPQSNCLKREKQDSCSITQALPSELKVSAD--NVSUTGAVSLASMLTE 241  
 Qy 249 IFLLQNAQAMPEVAVQRLKGAENWVSLSLHNAQFNLMKTPYIARHKGTPILLOQIDTAL 308  
 Db 242 IFLLQNAQAMPEVAVQRLKGAENWVSLSLHNAQFNLMKTPYIARHKGTPILLOQIDTAL 301  
 Qy 309 T---LQIDAQOQKLPISAGNRVFLFGGHDNTNIAAGMLGADWQLPQPDNTPPGGGLVF 365  
 Db 302 TPHPKQKQAYGVTLPTS-----VLFIAGHDTNLANLGGALELNWTLPGQPDNTPPGGGLVF 357  
 Qy 366 ELQNPNDHQRVYAVVMFYQTMQDLQNAEKLDLKNPAGIISVAVAGCE 414  
 Db 358 ERWRRLSDNSQWISQVSLVFQTLQOMDKTPLSL-NTPPGVKLTLAGCE 405

```
RESULT 15
US-10-034-985-2
; Sequence 2, Application US/10034985
; Publication No. US20030049815A1
; GENERAL INFORMATION:
; APPLICANT: Kretz, Keith
; TITLE OF INVENTION: NOVEL PHYTASE
; FILE REFERENCE: 09010/029003
; CURRENT APPLICATION NUMBER: US/10/034,985
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: US/09/580,515
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: 08/910,798
; PRIOR FILING DATE: 1997-08-13
; PRIOR APPLICATION NUMBER: 09/259,214
; PRIOR FILING DATE: 1999-03-01
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-034-985-2

Query Match      42.4%; Score 932.5; DB 14; Length 440;
Best Local Similarity 47.7%; Pred. No. 1.5e-82;
Matches 195; Conservative 61; Mismatches 138; Indels 15; Gaps 7;

Qy 13 LVLMISGLAATAPVA---APPSGYTLERVVILSRHGVRSPTKTOLMNDVTPDKWPMP 69
Db 5 LIPFLSLIPITPOSAPQSEPE-LKLESVVIVSRHGVRAPTKATQLMQDVTPDANETWP 63

Qy 70 VKAGYLTPRGAELVTLMGFGFYGYRSISGLIA-AGCPAEGGVAAQADIDQRTLTGQAF 128
Db 64 VKLWLTFRGGELIAYLGHYQRLVADGLLAKKGCPSQGVALLADVDERTRKTGEAFA 123

Qy 129 DGVAPGCGLTTHNQADLKKTDPLFHPVETGVCKLDNAQTDKAIERLGGPLDVTVSQRYAK 188
Db 124 AGLAPDCALTHTQADTSSPDPLFNPLKTVGCOLDNANVTDAILSRAGGSIADFTGHRQT 183

Qy 189 PFAQMGDVLFNAASPYCKSLQQCGKTCDFAHFAANEVNVNKEGTVKTLSPGLALSSTLGE 248
Db 184 AFRELERVLPQSNLCRKQDESCSLTQALPSELKVSAD--NVSLTGAVSLASMLTE 241

Qy 249 IFLQNAQMEVAVQRUKGAENVVLSLHNAQFNLMAKTPYIARHKGTPLQQIDTAL 308
Db 242 IFLQQAQMPGPGWGRITDSHOWNTLLSLHNAQFYLLIQTPEVARSRAPEPLDLINAA 301

Qy 309 T---LQLDAQGKLPISAQNEVFLFGHDTNIANIAGMLGADWOLPECPDNTPPGGGLVF 365
Db 302 TPHPQKQAYGVTLPTS----VLFAGHDTNLANLGGALELNTWLPQDPNTPPGGELVF 357

Qy 366 ELWQNPDNHQRVAVKMFYQTMQDLRNAEKLDLKXNPAGIISVAVAGCE 414
Db 358 ERWRRLSDNSQWIVQSLVFQTLQQRDKTPLSL-NTPEGEVKLTLAGCE 405
```

Search completed: April 30, 2004, 12:55:02  
Job time : 42.9756 secs

OM protein - protein search, using sw model

Run on: April 30, 2004, 12:35:05 ; Search time 15.3659 Seconds  
(without alignments)  
2629.235 Million cell updates/sec

Title: US-10-021-723B-4  
Perfect score: 2200  
Sequence: 1 MSVLENRVRLSGLVLMUSGL.....NPAGIISVAVAGCENNGDDK 420

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_78:.\*  
1: PIR1:.\*  
2: PIR2:.\*  
3: PIR3:.\*  
4: PIR4:.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2183	99.2	441	2 AC0201	acid phosphatase (
2	939.5	42.7	432	2 B36733	acid phosphatase (
3	926.5	42.1	434	2 H90770	phosphoanhydride p
4	926.5	42.1	444	2 D85633	hypothetical prote
5	532	24.2	413	2 F90773	periplasmic glucos
6	532	24.2	413	2 B85636	periplasmic glucos
7	530	24.1	413	2 JV0087	glucose-1-phosphat
8	504	22.9	414	2 E87316	periplasmic phosph
9	495	22.5	413	2 AG0632	glucose-1-phosphat
10	462	21.0	417	2 S25627	glucose-1-phosphat
11	125	5.7	411	2 D88504	protein B0361.7 li
12	124	5.6	421	2 S14742	acid phosphatase (
13	121.5	5.5	423	1 S06167	acid phosphatase (
14	121	5.5	438	2 S64682	acid phosphatase (
15	115.5	5.2	416	2 T16058	hypothetical prote
16	113	5.1	588	2 AF2542	hypothetical prote
17	108	4.9	26926	1 I38344	titin, cardiac mus
18	107.5	4.9	1940	2 F75393	hypothetical prote
19	106.5	4.8	541	2 S53799	nucleoporin NUP57
20	106	4.8	423	2 A33395	acid phosphatase (
21	105	4.8	468	2 A86233	hypothetical prote
22	103.5	4.7	505	1 P1WLRB	Li protein - cotto
23	102	4.6	537	2 S54770	secreted acid phos
24	102	4.6	888	2 T46726	secreted acid phos
25	101.5	4.6	479	2 F86535	6-phosphogluconate
26	101.5	4.6	479	2 C72088	6-phosphogluconate
27	101.5	4.6	715	2 AC0018	probable membrane
28	100.5	4.6	348	2 AB1751	B. subtilis CggR h
29	100.5	4.6	840	2 T02164	hypothetical prote

30 100 4.5 609 2 AF2379 proteinase IV [limp  
31 99.5 4.5 465 2 JE0369 histidine acid pho  
32 99.5 4.5 4367 1 B54802 dynein heavy chain  
33 99 4.5 440 2 T32457 hypothetical prote  
34 99 4.5 3525 2 E83641 probable hemagglut  
35 98.5 4.5 411 1 EDBEM4 45K immediate-earl  
36 98.5 4.5 586 2 F64186 ABC-type transport  
37 98.5 4.5 609 1 A42537 gene 16 protein -  
38 98 4.5 436 2 T16638 hypothetical prote  
39 97.5 4.4 348 2 AD1382 O-antigen biosynth  
40 97.5 4.4 1275 2 T18556 B. subtilis CggR h  
41 96.5 4.4 467 2 T04540 adenylosuccinate l  
42 96.5 4.4 527 2 T33175 hypothetical prote  
43 95 4.3 3705 2 AD0123 probable autotrans  
44 95 4.3 4485 2 T08044 dynein gamma heavy  
45 94.5 4.3 625 2 AC0309 phosphogluconate d

ALIGNMENTS

RESULT 1

AC0201  
acid phosphatase (EC 3.1.3.2) [imported] - Yersinia pestis (strain CO92)  
C:Species: Yersinia pestis  
C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AC0201  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.;  
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall  
Nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AC0201  
A>Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-441 <CUR>  
A:Cross-references: GB:AL590842; PIDN:CAC90470.1; PID:gl5979685; GSPDB:GN00175  
C:Genetics:  
A:Gene: YPO1648  
C:Keywords: phosphoric monoester hydrolase

Query Match	99.2%	Score	2183	DB	2	Length	441
Best Local Similarity	99.3%	Pred. No.	1.2e-168				
Matches	417	Conservative	1	Mismatches	2	Indels	0
Gaps	0						
Qy	1	MSVLENRVRLSGLVLMUSGLAATAPVAAPSPGYTLERVVILSRHGVRSPTKOTQLMNDV	60				
Db	1	MSVLENRVRLSGLVLMUSGLAATAPVAAPSPGYTLERVVILSRHGVRSPTKOTQLMNDV	60				
Qy	61	TPDKWPQWPVKAGYLTTPRGAEVLTLMGGFTYDIFRSGLLAAGCPAEGGYAADIQRT	120				
Db	61	TPDKWPQWPVKAGYLTTPRGAEVLTLMGGFTYDIFRSGLLAAGCPAEGGYAADIQRT	120				
Qy	121	RLTQQAFLDGVAPCGGLTVHNOADLKTDFLHPVETGVCKLDNAQTDKAIERLGGPLD	180				
Db	121	RLTQQAFLDGVAPCGGLTVHNOADLKTDFLHPVETGVCKLDNAQTDKAIERLGGPLD	180				
Qy	181	TVSORYAKPFAQMGDVLNFAASPYCKSLQOQKTCDFAPFAANEVNVNKGKTVTLGGPL	240				
Db	181	TVSORYAKPFAQMGDVLNFAASPYCKSLQOQKTCDFAPFAANEVNVNKGKTVTLGGPL	240				
Qy	241	ALSTLGEIFLLQNAQAMPVAVQRLKGAENWVLSLHNAQNLMAKTPYIAHKGTPL	300				
Db	241	ALSTLGEIFLLQNAQAMPVAVQRLKGAENWVLSLHNAQNLMAKTPYIAHKGTPL	300				
Qy	301	LQQIDTALTQLDQAQOGLPISAQNRVYFLGGHDTNTIANIAGMLGADWQPEQPDNTPPG	360				
Db	301	LQQIDTALTQLDQAQOGLPISAQNRVYFLGGHDTNTIANIAGMLGADWQPEQPDNTPPG	360				
Qy	361	GGLVFELWQNDHQRVAVKMFYQTDQLRNAEKLDKNNPAGIISVAVAGCENNGDDK	420				
Db	361	GGLVFELWQNDHQRVAVKMFYQTDQLRNAEKLDKNNPAGIISVAVAGCENNGDDK	420				





Db 243 EIFLLQQAQMPGCGRITDSHOWNTLLSLHNAQFYLLQRTPEVARSATPLLDLIMIA 302  
Qy 308 LT---LQIDAQOQKLPISAQRVFLGCHDNIANIAGMLGADWQLPQPDNTPPGGGLV 364  
Db 303 LTPHPQQAQVGVTLPTS-----VLFIAHDNLNLAGALELNTLTPQPDNTPPGGGLV 358  
Qy 365 FELWNPNDHORYAVKVFYQTMQDRNAEKLKNNPAGIISVAVAGCE 414  
Db 359 FERWERLSDNSQWIQVSLVFTLQQRDKTFLSL-NTPPGEVKLTLAGCE 407  
RESULT 4  
D85633  
hypothetical protein appA [imported] - Escherichia coli (strain O157:H7, substrain EDL933)  
C:Species: Escherichia coli  
C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C:Accession: D85633  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: D85633  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-444 <STO>  
A:Cross-references: GB:AB005174; NID:G12514245; PIDN:AAG55528.1; GSPDB:GN00145; UWGP:Z13  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: appA

Query Match 42.1%; Score 926.5; DB 2; Length 444;  
Best Local Similarity 47.3%; Pred. No. 5e-67;  
Matches 194; Conservative 62; Mismatches 139; Indels 15; Gaps 7;  
Qy 13 LVLMGLSLAAITAPVA---ASPS-CYTLERVVLSRHGVRSPTKOTOLMNDVTPDKWPQW 68  
Db 15 LIPFLSLIPLTPQSAPAQSEPELKLSEVIVSRHGVRAPTKATQLMQDVTDPANPW 74  
Qy 69 PVKAGYLTFRGAEVLTMGGFYGYDYSRLGLLA-AGCPAEGGVYAQADIDORTLTGQAF 127  
Db 75 PVKGLWLTFRGAEVLTMGGFYGYDYSRLGLLA-AGCPAEGGVYAQADIDORTLTGQAF 134  
Qy 128 LDGVAPGCGLTVHQAQDLKKTDPFLPHVETGVCKLDNAQTDKAIERLGGPLDTSQRYA 187  
Db 135 AAGLAPDCAITVHQADTSPPDLPFNLTGVVGLDNNVTDALLSRAGGSIADFTGHRQ 194  
Qy 188 KPFAQMGDVLNFAASPYCKSLQOQKTCDFAHFAANVNNKGTGKVTLSGLSALSTLG 247  
Db 195 TAFRELRLVNFQSNCLNREKQDESCSLQALPSELKVSAD--NVSLTGAVALSMLT 252  
Qy 248 EIFLLQQAQMPGCGRITDSHOWNTLLSLHNAQFYLLQRTPEVARSATPLLDLIMIA 302  
Db 253 EIFLLQQAQMPGCGRITDSHOWNTLLSLHNAQFYLLQRTPEVARSATPLLDLIMIA 312  
Qy 308 LT---LQIDAQOQKLPISAQRVFLGCHDNIANIAGMLGADWQLPQPDNTPPGGGLV 364  
Db 313 LTPHPQQAQVGVTLPTS-----VLFIAHDNLNLAGALELNTLTPQPDNTPPGGGLV 368  
Qy 365 FELWNPNDHORYAVKVFYQTMQDRNAEKLKNNPAGIISVAVAGCE 414  
Db 369 FERWERLSDNSQWIQVSLVFTLQQRDKTFLSL-NTPPGEVKLTLAGCE 417

RESULT 5  
F90773  
periplasmic glucose-1-phosphatase [imported] - Escherichia coli (strain O157:H7, substra  
C:Species: Escherichia coli  
C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
C:Accession: F90773  
R:Havashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and ge  
A:Reference number: A99629; MUID:21156231; PMID:11238796  
A:Accession: F90773  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-413 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA834581.1; PID:G13360618; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RMD 0509952  
C:Genetics:  
A:Gene: EC61158

Query Match 24.2%; Score 532; DB 2; Length 413;  
Best Local Similarity 31.3%; Pred. No. 3.6e-35;  
Matches 131; Conservative 83; Mismatches 164; Indels 40; Gaps 11;  
Qy 10 LSGVLMLSLGAAITAPVAAPSGYTLERVLVLSRHGVRSPTKOTQLMNDVTPDKWPQW 68  
Db 10 VAGVLLASNAQAQTV-----PEGVLQQLVLMMSRHLNRLAPLANNNGSVLEQSTPNKWPW 64  
Qy 69 PVKAGYLTFRGAEVLTMGGFYGYDYSRLGLLAAG-CPAEGGVYAQADIDORTLTGQAF 127  
Db 65 DVPGGQLTKGGVLEVMGHYMRLEAQGMVKSGECPDPTVYAYANSQRTVATAQPF 124  
Qy 128 LDGVAPGCGLTVHQAQDLKKTDPFLPHVETGVCKLDNAQTDKAIERLGGPLDTSQRYA 187  
Db 125 ITGAFFGCDIPVHHQSKGTMDPTFNPTVITDSSAFAFSEQAVAMEKEUS-----KLQT 178  
Qy 188 KPFAQMGDVLNFAASPYCKSLQO---QKTCDFAHFAANVNNKGTGKVTLSGLSAL 243  
Db 179 DSVQLLEKIVNKKDSPAECQKQCSLVGDKNT---FSA-----KYQEPGVSGFLKVG 228  
Qy 244 STLGEIFLQNAQAMP--EVANQRLKGAENWVLSLHNAQFYLLQRTPEVARSATPLLDLIMIA 301  
Db 229 NSLVDAFTLQYVGGPMDQVANGETKSDQKWKSLKNGYQDSLFTSEVARNVAKPLV 288  
Qy 302 QCIDTALTLDQAQOQKLPISAQRVFLGCHDNIANIAGMLG-ADWOLPQPDNTPPG 360  
Db 289 SYIDKALVTR-----TSAPKITVLVGHDSNIALSLTALDFFKYLQHDQERTPTG 339  
Qy 361 GGLVPELWNPNDHORYAVKVFYQTMQDRNAEKLKNNPAGIISVAVAGCE 418  
Db 340 GKIVFORWRDSKANRLMKIEYVQSAEQLRNADALTQ-APAQRVLTLESLGCPDAD 396

RESULT 6  
B85636  
periplasmic glucose-1-phosphatase [imported] - Escherichia coli (strain O157:H7, subst  
C:Species: Escherichia coli  
C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C:Accession: B85636  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayh  
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodac  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: B85636  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-413 <STO>  
A:Cross-references: GB:AB005174; NID:G12514273; PIDN:AAG55550.1; GSPDB:GN00145; UWGP:Z  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: agp

Query Match 24.2%; Score 532; DB 2; Length 413;  
Best Local Similarity 31.3%; Pred. No. 3.6e-35;  
Matches 131; Conservative 83; Mismatches 164; Indels 40; Gaps 11;  
Qy 10 LSGVLMLSLGAAITAPVAAPSGYTLERVLVLSRHGVRSPTKOTQLMNDVTPDKWPQW 68  
Db 10 VAGVLLASNAQAQTV-----PEGVLQQLVLMMSRHLNRLAPLANNNGSVLEQSTPNKWPW 64  
Qy 69 PVKAGYLTFRGAEVLTMGGFYGYDYSRLGLLAAG-CPAEGGVYAQADIDORTLTGQAF 127

Db 65 DVPGGQLTTKGGVLEVMGHMREWLAEQGWKSGECPPDPDTYYAVANSIQRTVATAQFF 124  
Qy 128 LDGVAFCGLTVNQADLKTDPFLPHVETGVCKLDNAQTDKAIERLGGPLDVTVSQRYA 187  
Db 125 ITGAFPGCDIPVHQEKMGKMTDFNPTVITDDSAFSEQAVAAAMEKELS-----KLQLT 178  
Qy 188 KPFAQMGDVLFNFAASPYCKSLQO-----QKTKCDFAHFAANVNNKEGTVKLSGFLALS 243  
Db 179 DSYQLEKIVNYKDSPACKERQOCSLVDGKNT-----FSA-----KYQEPGVSGPLKV 228  
Qy 244 STLGEIFLLQNAQAMP--EVAWQRLKGAENWVLSLHNAQFNLMKTPVIARHKGTPLL 301  
Db 179 DSYQLEKIVNYKDSPACKERQOCSLVDGKNT-----FSA-----KYQEPGVSGPLKV 228  
Qy 244 STLGEIFLLQNAQAMP--EVAWQRLKGAENWVLSLHNAQFNLMKTPVIARHKGTPLL 301  
Db 229 NSLVDATFLQYEGFFMDQVANGKISDQKQVLSKLKNGYQDSLFTSPVARNVAKPLV 288  
Qy 302 QCIDTALTQLDQAQOKLPISAQNRVFLFGCHDTNANIAGMLG-ADWOLPEQPDNTPPG 360  
Db 289 SYIDKALVTR-----TSAPKITVLVGHDSNTASLTALDFKPYQLHDQNERTPIG 339  
Qy 361 GGLVFEWQNDHQRVAVVMFYQTMQDLNRNAEKLDKNNPAGIISVAVAGCENNGD 418  
Db 340 GKIVFORWRDSKANRDLMKIEVYQSAEQURNADALTQ-APAQRVTLELSCGPIDAD 396

## RESULT 7

JV0087

Glucose-1-phosphatase (EC 3.1.3.10) precursor - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C&gt;Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 01-Mar-2002

C:Accession: J00087; H64841

J:Pradel, B.; Marck, C.; Boquet, P.L.

J: Bacteriol. 172, 802-807, 1990

A:Title: Nucleotide sequence and transcriptional analysis of the Escherichia coli agp gene

A:Reference number: JV0087; MUID:90130318; PMID:2153660

A:Accession: J00087

A:Molecule type: DNA

A:Residues: 1-413 &lt;PRA&gt;

A:CROSS-references: GB:M33807; NID:G145217; PIDN:AAA23426.1; PID:G145218

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, D.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: H64841

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-413 &lt;BLAT&gt;

A:CROSS-references: GB:U00096; NID:G1787233; PIDN:AACT4087.1; PID:G1787237;

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: agp

C:Function:

A:Map position: 23 min

A:Description: essential for growth in a high-phosphate medium containing glucose-1-phosphate

A:Note: optimal at low pH

A:Keywords: homodimer; periplasmic space; phosphohistidine; phosphoprotein; phosphoric acid

F:1-22/Domain: signal sequence #status predicted &lt;SIG&gt;

F:23-413/Product: glucose-1-phosphatase #status predicted &lt;MAT&gt;

F:40/Active site: His (phosphohistidine intermediate) #status predicted

F:311/Active site: His #status predicted

Query Match 24.1%; Score 530; DB 2; Length 413;

Best Local Similarity 31.3%; Pred. No. 5.2e-35;

Matches 131; Conservative 82; Mismatches 165; Indels 40; Gaps 11;

Qy 10 LSGVLVLSGLIAITAPVAAPSGYTLERVVLSRHGVRSP-TKQTQMDNDVTPDKWPOW 68

Db 10 VAGIVLLASNAQAQTV-----PEGYQLQVVLMSRHLNRAFLANNNGSVLQSTPNKWEW 64

Qy 69 PVKAGYLTTPGAEVLTLMGFGYDYSRLGGLLAAG-CPAEGGVYQAQADIDQRTLTGQAF 127

Db 65 DVPGGQLTTKGGVLEVMGHMREWLAEQGWKSGECPPDPDTYYAVANSIQRTVATAQFF 124

Qy 128 LDGVAFCGLTVNQADLKTDPFLPHVETGVCKLDNAQTDKAIERLGGPLDVTVSQRYA 187  
Db 125 ITGAFPGCDIPVHQEKMGKMTDFNPTVITDDSAFSEQAVAAAMEKELS-----KLQLT 178  
Qy 188 KPFAQMGDVLFNFAASPYCKSLQO-----QKTKCDFAHFAANVNNKEGTVKLSGFLALS 243  
Db 179 DSYQLEKIVNYKDSPACKERQOCSLVDGKNT-----FSA-----KYQEPGVSGPLKV 228  
Qy 244 STLGEIFLLQNAQAMP--EVAWQRLKGAENWVLSLHNAQFNLMKTPVIARHKGTPLL 301  
Db 229 NSLVDATFLQYEGFFMDQVANGKISDQKQVLSKLKNGYQDSLFTSPVARNVAKPLV 288  
Qy 302 QCIDTALTQLDQAQOKLPISAQNRVFLFGCHDTNANIAGMLG-ADWOLPEQPDNTPPG 360  
Db 289 SYIDKALVTR-----TSAPKITVLVGHDSNTASLTALDFKPYQLHDQNERTPIG 339  
Qy 361 GGLVFEWQNDHQRVAVVMFYQTMQDLNRNAEKLDKNNPAGIISVAVAGCENNGD 418  
Db 340 GKIVFORWRDSKANRDLMKIEVYQSAEQURNADALTQ-APAQRVTLELSCGPIDAD 396

## RESULT 8

E87316

periplasmic phosphoanhydride phosphohydrolase [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C&gt;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001

C:Accession: E87316

R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, B.; Iamb, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Hatt, D.H.; Kol

n, J.; Emlaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: E87316

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-414 &lt;STO&gt;

A:CROSS-references: GB:AE005673; NID:G13421731; PIDN:AAK22529.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC0542

Query Match 22.9%; Score 504; DB 2; Length 414;

Best Local Similarity 35.0%; Pred. No. 6.7e-33;

Matches 140; Conservative 65; Mismatches 171; Indels 24; Gaps 14;

Qy 20 LAATAPVAA--EPGVTLEVRVILSRHGVSRPTKQTQLMNDVTPDKWQPVKAGYLT 77

Db 7 IAVVTLALASGASAELEKLVILSRHGVSRPMSSEERLEASARFPRFVAGHLTA 66

Qy 78 RGAELVTLMGGFYGDYFRSLGLLAAG-CPAEGGVYQAQADIDQRTLTGQAFLDGVAPCG 136

Db 67 RGETLVARNGDYRRHYAAQGLLXPGDC---ASVYAWANVTQRTIATAKAYRETLPAGCP 123

Qy 137 LTVNQADLKTDPFLPHVETGVCKLDNAQTDKAIERLGGPLDVTVSQRYAKPFAQMGDV 196

Db 124 VTV-NTVNGNIDPFEPVKGIVKADHALARAAGVGGDLTAWASHNQAEQLDAL 182

Qy 197 LNFAASPYCKSLQOQKTKCDFAHFAANVNNKEGTVKLSGFLALSSTLGEIFLL--QN 254

Db 183 LMQCDKGPAPGKRRV-----FDAKPGFVDGE-ELAGLSGPEAFASGVTESILMAWAD 236

Qy 255 AQMPVAVQRLKGAENWVLSLHNAQFNLMKTPVIARHKGTPFLQCIDTALTQLDA 314

Db 237 GRDFAGLQKWSL-DEEALTRSFLLHQAEFFDLRLTPPVARTLHGLADRL--AATLR-DG 292

Qy 315 CQOKLPISAQNRVFLFGCHDTNANIAGMLGADQLP-EQPDNTPPGGLVFEWQNDP 373

Db 293 AAAIGFVDA--RLVIAIGHDGTASLGLLRLMENTLPQYQPNQIQPGGALVFERWR--DD 349

Qy 374 HORYVAVVMFYQTMQDLNRNAEKLDKNNPAGIISVAVAGC 413

Db 350 GYRVVRVFTQGSLSQLRNMTALDAKTPPLS-APVFGQC 388

RESULT 9

AG0632  
Glucose-1-phosphatase precursor (Glpase), secreted [imported] - Salmonella enterica subsp. C:Species: Salmonella enterica subsp. enterica serovar Typhi  
A:Note: this species has also been called Salmonella typhi  
C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C:Accession: AG0632  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Comerford, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Garra, P.  
Nature 413, 848-852, 2001  
A:Authors: Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Comerford, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Garra, P.  
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi  
A:Reference number: AB0502; MUID:21534947; PMID:11677608  
A:Accession: AG0632  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-413 <PAR>  
A:Cross-references: GB:AL513382; PIDN:CAD08242.1; PID:G16502289; GSPDB:GN00176  
C:Genetics:  
A:Gene: STY1153

Query Match 22.5%; Score 495; DB 2; Length 413;  
Best Local Similarity 31.8%; Pred. No. 3.6e-32;  
Matches 134; Conservative 78; Mismatches 166; Indels 44; Gaps 14;

QY 8 VRLSLGLVLMGLAATAAPVAAPSGYTLERVILSRHGVRSPTKQTLQMDVTPDKWP 66  
DB 8 VAVAGAVLLSSAVQAQTT-----PEGYLOQVLMWSRHLAPLANNGVLAQSTPNWP 62  
QY 67 QNPVAGYILTPRGABLVLMGGFYGDYFRSLGLAAG-CPAEGGVYAAQADIDQRTLTGQ 125  
DB 63 AWDVPGGLTTKGGVLYVMGHTREMLVAQGLIPSGCPADPTVYAVANSLOQTATAQ 122  
QY 126 AFLDGVAPCGLTJHQAADLKTDPLFHPVEVGVCKLQNAQTDKAIEERLGGPLDVTVSOR 185  
DB 123 FITSAFGCDIPVHQKMGTMDFTPVITDSSAFAQVQWEX-----ARSQL 175  
QY 186 YA-KPFAQMGVDVNFAPSPYCKSLQOQKTCDFAH----FAANEVNVNKEGTKVTLSGPL 240  
DB 176 HLDESQYKLEQITHYQDSFSCKEHQ----CSLIDAKDTFSANTYQ-----EPGVQGLP 225  
QY 241 ALSSLTGLIFLQNAQAMP--EVAWQLKGAENWVLSLHNAQFNLMKTPYARHKG 298  
DB 226 KVGNSLDAFTLQYIEGFPMDQVAMGGHTDRQMKVLSKLKNGYQDSLFTPTVARNVA 285  
QY 299 FLQOQIDTALTQLDAQOQKLPISAQNRVFLGGHDTNANIAGMLG-ADWQLPEQPDNT 357  
DB 286 PLVKYIDKVLVAE-----RVSAAP-KVTVLVGHDSNLSILTALDFKPYQLHDQYERT 336  
QY 358 PPGGLVFWLQNPNDHQRVAVVMFYQTMQDLRNAEKLKNNPAGIISVAVAGC--EN 415  
DB 337 PIGGQLVFQWHDGNANRLMKIEVYQSAQRLRNAEALTLK-SPAQRVTLKLGCPVDA 395  
QY 416 NG 417  
DB 396 NG 397

RESULT 10

S25627  
Glucose-1-phosphatase (EC 3.1.3.10) precursor - Providencia rettgeri  
C:Species: Providencia rettgeri  
C:Date: 25-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 08-Oct-1999  
C:Accession: S25627  
R:Riccio, M.B.; Chaturvedi, A.; Lombardi, G.; Satta, G.  
submitted to the EMBL Data Library, September 1992  
A:Reference number: S25627  
A:Accession: S25627  
A:Molecule type: DNA  
A:Residues: 1-417 <RIC>  
A:Cross-references: EMBL:X68201; NID:G45771; PIDN:CAA48288.1; PID:G45772

A:Experimental source: strain PV7

C:Genetics:  
A:Gene: agp  
C:Keywords: periplasmic space; phosphohistidine; phosphoprotein; phosphoric monoester  
F:1-29/Domain: signal sequence #status predicted <SIG>  
F:30-417/Product: Glucose-1-phosphatase #status predicted <WAT>  
F:43/Active site: His (phosphohistidine intermediate) #status predicted  
F:314/Active site: His #status predicted

Query Match 21.0%; Score 462; DB 2; Length 417;  
Best Local Similarity 30.6%; Pred. No. 1.7e-29;  
Matches 128; Conservative 84; Mismatches 170; Indels 36; Gaps 15;

QY 14 VLMGLSLAATAAPVAAPSGYTLERVILSRHGVRSPTKQTLQMDVTPDKWP 67  
DB 5 VLTLCLSALFAAPPTMASTNQADMVLDQVLSRHLRTPVNTGTLTETDKKND 64  
QY 68 WPKVAGYILTPRGABLVLMGGFYGDYFRSLGLAAG-CPAEG-GVYAAQADIDQRTLTGQ 125  
DB 65 WDAKSYLTQGCALSVVMGHYFREWIDQNKLLADELCPTSNEDIYVLTNSLQRTATAQ 124  
QY 126 AFLDGVAPCGLTJHQAADLKTDPLFHPVEVGVCKLQNAQTDKAIEERLGGPLDVTVSOR 185  
DB 125 FFAAGAFPGCKVNIHQPIGKMDPVFNPIITNGSPFKQKALAMDDYLKG----LSUK 180  
QY 186 YAKPFAQMGVDVNFAPSPYCKSLQOQKTCDF--AHFAANEVNVNKEGTKVTLSGPLSS 244  
DB 181 AG--YEELDTVLNIDSKCKT----DKLCLNDSQKNSFIIEADKE---PGVSGPLKIAN 231  
QY 245 TLGSIFFLQNAQAMP--EVAWQLKGAENWVLSLHNAQFNLMKTPYARHKGTPILQ 302  
DB 232 SAYDAIDLOYYEGFPADQVAVGLVDTPEKWKLTNKNAYQBTLEFPKIIAKNVAPILN 291  
QY 303 QIDTALTQLDAQOQKLPISAQNRVFLGGHDTNANI--AGMLGADWQLPEQPDNTPPGG 361  
DB 292 YIDKGF-VSVD-KGETA-----KTFVLVGHDSNLSIASMSAMDFKPYQLAQYEHPTIGG 343  
QY 362 GLVFELWQNPNDHQRVAVVMFYQTMQDLRNAEKLKNNPAGIISVAVAGC--ENNG 417  
DB 344 KLVFORWTDQTKKCFMKVEYVYQADQLRDNAYLSLETPPKH-VTLELKDQCPVDKNG 400

RESULT 11

D88504  
Protein B0361.7 [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 24-May-2001  
C:Accession: D88504  
R:anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biol  
A:Reference number: A75000; MUID:99065613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C\_e  
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; a  
A:Accession: D88504  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-411 <STO>  
A:Cross-references: GB:chr\_III; PIDN:AAA50626.1; PID:G458955; GSPDB:GN00021; CESP:B036  
C:Genetics:  
A:Gene: B0361.7  
A:Map position: 3  
C:Superfamily: mammalian acid phosphatase

Query Match 5.7%; Score 125; DB 2; Length 411;  
Best Local Similarity 20.5%; Pred. No. 0.03;  
Matches 91; Conservative 58; Mismatches 161; Indels 134; Gaps 20;

QY 13 LVLMGLSLAATA-----PVAAPSGYTLERVILSRHGVRSPTKQTLQMDVTPDKWP 66  
DB 4 LVLVLLGASGINAVVYKEVPIQANTD--TLEYVHTVWRHGDRTFAEL-----LFDDIT 55  
QY 67 QNPVAGYILTPRGAE-----LVLMGGFYGDYFRSLGLAAGCPAEGGVYAAQADIDQ 118

[illegible]

## RESULT 12

S14742  
 Acid phosphatase (EC 3.1.3.2) precursor - mouse (fragment)  
 CSpecies: Mus musculus (house mouse)  
 CDate: 21-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 31-Mar-2000  
 CAccession: S14742  
 R:Geiger, C.; von Figura, K.; Pohlmann, R.  
 Biol. Chem. Hoppe-Seyler 372, 301-304, 1991  
 ATitle: Molecular cloning of the mouse lysosomal acid phosphatase.  
 AReference number: S14742; MUID:91282986; PMID:2059337  
 AAccession: S14742  
 AStatus: preliminary  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-421 <GEI>  
 A:Cross-references: EMBL:X57199; NID:G52870; PIDN:CAA40485.1; PID:G52871  
 C:Superfamily: mammalian acid phosphatase  
 C:Keywords: lysosome; phosphoric monoester hydrolase

Query Match	5.6%;	Score 124;	DB 2;	Length 421;
Best Local Similarity	22.8%;	Pred. No. 0.037;		
Matches 100;	Conservative 52;	Mismatches 148;	Indels 138;	Gaps 25;
QY	11	SGVLMLSLGA-AITAPVAAEPGVTILVERVILSRHGVRSPTRKOTQLMNDVTPDKWPQPW 69		
DB	9	AALLQFLQWCLUTVNPPIQAR-----SLRFVTLVLRHGRDSFVK-TPKPGPYQEEKHPQ-- 61		
QY	70	VKAGVLTIPRGA-----EL-VTLMGGFYG-----DYFRSLGLLAAQCPAEGGYAAQADIDQRT 120		
DB	62	-GFGQLTKEGMLQHWELGQALRQRVHGFLNTSYHRQ-----EVVVRSTDFDT 108		
QY	121	RUTGQAFLDGVAPEGGLTVHMQADLKTKTDPLFHPVETGVCKLDNAQTDKAIERLGGPLD 180		
DB	109	LMSAEANLAGLFP-----PNEVQHFSENIISWQIP-----VHTVPTIEDRL----- 149		
QY	181	TWSQRVAKFPAQMGVDNLNFAASPYCKSLQQQCKT-----CDFAHFAAEVNV 227		
DB	150	-----LKPLGFCPRYEQLNQETRONPEYQNRISIQNAQPLNVAVNETGL 193		
QY	228	NKEGTKVTLSGGLALSSLTGEIF-----LLQNAQAMPEVAWORLGAEB--NWVSL 275		
DB	194	-----TNVTL-----EITNNVYDTLFCEQTHGULLPPWASPQIV-QRLSQLKDFSLFL 241		
QY	276	LSLHNAQFNLMAKTPYIAHKGTPTLLQIDTALTALTQLDAQCKLPISAQNRVFLFLGHGDT 335		
DB	242	FGTH-----EVOVKARLOGCVLLAAILKNLITM--ATTSOPF-----KLIVYSAHDT 286		

Db 152 -----LAFPLGCPRYEQLNQETQTPYQNESSRNAQFLDMVANE 192  
QY 225 VVNKGEKTVTLGSLSSSTLGEIFLLQNAQAPPEVAVQRLKGAENWVSLSLHNAQEN 284  
Db 193 TGL-----TDLTETVWNYDTL-----FCEQTHGLRLPPW-----ASPQWQRSLRUKPSFR 241  
QY 285 LM---AKTPYIARHKGTPLLQIDTALTALQDAQCKLPIISAQNRVFLGLGHDNTNANIA 341  
Db 242 FLFGIYQQAERKARLGGVLLAQIRKGLTM-ATTSQLP-----KLLVYSAHDTTL--VA 292  
QY 342 GMLGADWQLPPQDNTPTPGGLVFLWQNPNDHQRVAVKMFYQT----- 386  
Db 293 LQMALDVYNGEQ---APYASCHIFELYQEDSGN---FSVEMYFRNESKAPWPLSLPGCP 346  
QY 387 -----MDOLRNAE-----KLDLKNPAG---IISVAVAG 412  
Db 347 HRCPLQDFLRLTEPVVPKDWQEQCLASGPADTEVIVALVAGC 389

RESULT 14  
S64682  
acid phosphatase (EC 3.1.3.2) Acph-1 precursor, lysosomal - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C>Date: 06-Dec-1996 #sequence\_revision 06-Dec-1996 #text\_change 17-Mar-2000  
C:Accession: S64682; S64681  
R:Chung, H.J.; Shafer, C.; MacIntyre, R.  
Mol. Gen. Genet. 250, 635-646, 1996  
A:Title: Molecular characterization of the lysosomal acid phosphatase from Drosophila melanogaster  
A:Reference number: S64681; MUID:96194627; PMID:8676866  
A:Accession: S64682  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-438 <CHU>  
A:Accession: S64681  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-31, 'Y', 33-438 <CHW>  
C:Genetics:  
A:Gene: FlyBase:Acph-1  
A:Cross-references: FlyBase:FBgn0000032  
C:Superfamily: mammalian acid phosphatase  
C:Keywords: phosphoric monoester hydrolase  
F:1-33/Domain: signal sequence #status predicted <SIG>  
F:34-438/Product: acid phosphatase Acph-1 #status predicted <MAT>

Query Match 5.5%; Score 121; DB 2; Length 438;  
Best Local Similarity 21.3%; Pred. No. 0.066;  
Matches 84; Conservative 60; Mismatches 151; Indels 100; Gaps 19;  
QY 36 LERVVILSRHGVSPKQTLQMDVTPDKWQO---WPVKAGYLTIRGAEVLTLMGGFYCD 92  
Db 48 LKPVHVIYRHGDRTP-----VDPYPTDFGDRKFTWTCWGDITNLGKQEHYDLGKWLRLN 101  
QY 93 YFSLGLLAAGCP---AEGGVYQA-DIDQRTLTGQAFLDGVPAGCGLTWHNQADLTKT 148  
Db 102 RYSNL-----LPPIYSNENIYVQSTDVD-RTLMSAQSNLAGLYEPQG-----EDWNT 148  
QY 149 DPLFHPVETGVCKLDNAQTDKATEERLGGPLDTVSQRYAKPFAQMG-----DVNLFAA 201  
Db 149 DINWQPI-----PIHTSPREDPILAAKAPCPAYDELASLES 186  
QY 202 SPYCKSLQOQKTCDFAHFAANVNVNKGTV-TLSGFLALSSLTGEIFLLQNAQAMPE 260  
Db 187 SPEFKALTEKHNLFAYLS-----EKGRPVKTFIDAQYLNNTL-----FIENLYNMTL 235  
QY 261 VAM-QRLKGAENWVSLSLHNAQFNLMKTPYIARHKGTPLLQIDTALTALQDAQCKL 319  
Db 236 PKWTKKVGRE---ELTYVSNFAFALISSYTRKLARLAKGFLKDI-----FORFKKSSG 287  
QY 320 PISAQNRVFLGLGHDNTNANIAQMLGADWQLPEQDNTPTPGGLVFLWQNPNDHQRVAV 379  
Db 288 SLKPRSRMWYSADDTTVAASVLNAL-----KLFLHSP-----YTA 324

QY 380 VKNFYQTMQDLRN-AEKLDLKNPAGIISVAVAGC 413  
Db 325 CIMMELRVDBTNTPLVSIYFYKNTTAEPLDLDPGC 359  
RESULT 15  
T16058  
hypothetical protein F13D11.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 23-Mar-2001  
C:Accession: T16058  
R:Fulton, L.  
submitted to the EMBL Data Library, November 1995  
A:Description: The sequence of C. elegans cosmid F13D11.  
A:Reference number: S69020  
A:Accession: T16058  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-416 <FUL>  
A:Cross-references: EMBL:U40939; NID:G1072175; PID:G1072178; PIDN:AAA81702.1; CESP:F13  
C:Genetics:  
A:Gene: CESP:F13D11.1  
A:Introns: 23/3; 59/3; 88/3; 138/3; 202/3; 244/1; 270/3; 312/3; 349/3; 377/1  
C:Superfamily: mammalian acid phosphatase  
Query Match 5.2%; Score 115.5; DB 2; Length 416;  
Best Local Similarity 21.6%; Pred. No. 0.18;  
Matches 90; Conservative 60; Mismatches 118; Indels 149; Gaps 25;  
QY 13 LVLMLSGLAITAPVAAPSPGYTLERVILSRHGVSPKQTLQ-----MNDVTPDK--WP 66  
Db 3 LVLMLPFL-----FPVAFG---KLKPVQIWFHRTGPHYLYFPDGLNNVDYQQIAMP 54  
QY 67 QWPVKAGYLTIRGAEVLTLMGGFYCDYFSLGLAAGCPAEGGVYQAADI----- 116  
Db 55 -----GELTKRGILBEFQLGQELRKIYGEHF-----GDTYQPRDFHYVTGK 95  
QY 117 DQRTLTGQAFLDGVPAGCGLTWHNQADLTKTDLFHPVETGVCKLDNAQ-TDKATEERL 175  
Db 96 DNRTSASQAQAMPAGFLP-----PNEDQTNWYELKWPV-----AQLTDESIDWVS 140  
QY 176 GGFLDTV-----SQRYAKPFAQMGD-----VLNFAASPYCKSLQOQKTCDFPA 218  
Db 141 LGAIDNCPVYGRAQKRSSEYAEVMDQEKYDABELLQVRNHADEPIVEAVK-----YN 193  
QY 219 HFAANVNVNKGTVTLGSLFALSSLTGEIFLLQNAQAMPEVAVQRLKGAENWVSLLS- 277  
Db 194 H-----VIDSLKVRVILQD-DRLFPPEWAR--GYENRIINMSF 228  
QY 278 -LHNAQFNLMKTPYIARHKGTPLLQIDTALTALQDAQCKLPIISAQNRVFLGLGHDNTN 336  
Db 229 LIHDAV--VKVQDSVGDYHNLVMSYFETHL-----OK-----NSTKGVFISGHDNTN 274  
QY 337 IANTAGMLGADWQ---LPEQDNTPTPGGLVFLWQNPNDHQRVAVKMFYQTMQDL 390  
Db 275 LVTI-----WESLRIDGHPEIDP-----NYGAHIALEM-HEPVGQL 309

Search completed: April 30, 2004, 12:42:18  
Job time : 18.3659 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 12:31:10 ; Search time 10.2439 Seconds  
(without alignments)  
2134.876 Million cell updates/sec

Title: US-10-021-723B-4  
Perfect score: 2200  
Sequence: 1 MSVLENRVRLSGLVLMISGL.....NPAGIISVAVACENNGDDX 420

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues  
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Length	ID	Description
1	939.5	42.7	432	1 PPA_ECOLI	P07102 escherichia
2	530	24.1	413	1 AGP_ECOLI	P19926 escherichia
3	499	22.7	413	1 AGP_SALTY	Q03921 salmonella
4	462	21.0	417	1 AGP_PRORE	Q52309 providencia
5	125	5.7	422	1 PPAY_CABEL	Q10944 caenorhabdi
6	124	5.6	423	1 PPAL_MOUSE	P24638 mus musculu
7	121.5	5.5	423	1 PPAL_HUMAN	P11117 homo sapien
8	106.5	4.8	541	1 NU57_YEAST	P48837 saccharomyc
9	106	4.8	423	1 PPAL_RAT	P20611 rattus norv
10	103.5	4.7	505	1 VLI_CRPVK	P03102 cottontail
11	101.5	4.6	479	1 6PGD_CHLEN	Q92813 chlamydia p
12	101.5	4.6	715	1 IGAA_YERPE	P58722 yersinia pe
13	99.5	4.5	4367	1 DVHC_NEUCR	P45443 neurospora
14	98.5	4.5	411	1 VIB2_HCMVA	P19893 human cytom
15	98.5	4.5	586	1 CYDD_HAEIN	P45082 haemophilus
16	98.5	4.5	609	1 VGI6_BPP22	Q01146 bacterioph
17	97.5	4.4	1275	1 RFSC_MYXXA	Q50864 myxococcus
18	95	4.3	743	1 PURL_RHIME	Q92977 rhizobium m
19	95	4.3	4485	1 DYHG_CHALRE	Q39575 chlamydomon
20	94.5	4.3	1049	1 EF3_CANAL	P25997 candida alb
21	93.5	4.2	1584	1 KYK1_DICDI	P18160 dictyosteli
22	92.5	4.2	410	1 VIB2_HCMVT	P06435 human cytom
23	92	4.2	513	1 C752_SOLME	P37120 solanum mel
24	91.5	4.2	666	1 FPCK_WAIZE	Q95120 zea mays (m
25	91.5	4.2	720	1 FPTA_PSEAE	P42512 pseudomonas
26	91.5	4.2	772	1 TLK2_HUMAN	Q86ue8 homo sapien
27	91	4.1	390	1 MALY_ECOLI	P23256 escherichia
28	91	4.1	487	1 MURE_PSESM	Q87wy0 pseudomonas
29	91	4.1	506	1 C751_PETHY	P48418 petunia hyb
30	91	4.1	745	1 P021_EIG	Q29076 sus scrofa
31	90.5	4.1	508	1 PSBB_EUGGR	P14813 euglena gra
32	90.5	4.1	1063	1 HGPI_HABIN	P44795 haemophilus
33	90	4.1	381	1 PPAP_RAT	P20646 rattus norv

Q8ywf0 anabaena sp  
Q8euy3 mycoplasma  
Q08336 cercocobus  
P55540 rhizobium s  
P52732 homo sapien  
Q03132 saccharopol  
P48419 petunia hyb  
P08489 serratia ma  
Q55669 synecocyst  
P00545 feline sarc  
P79184 macaca fusc  
Q08340 macaca neme

ALIGNMENTS

RESULT 1				
ID	PPA_ECOLI	STANDARD;	PRT;	432 AA.
AC	P07102;			
DT	01-APR-1988 (Rel. 07, Created)			
DT	01-AUG-1991 (Rel. 19, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Periplasmic appa protein precursor [Includes: Phosphoanhydride			
DE	phosphohydrolase (EC 3.1.3.2) (pH 2.5 acid phosphatase) (AP); 4-			
DE	phytase (EC 3.1.3.26)]			
GN	APPA OR B0980			
OS	Escherichia coli.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Escherichia.			
OX	NCBI_TaxID=562;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 23-29.			
RC	STRAIN=K12;			
RX	MEDLINE=90368616; PubMed=2168385;			
RA	Dassa J., Marck C., Boquet F.L.;			
RT	"The complete nucleotide sequence of the Escherichia coli gene appa			
RT	reveals significant homology between pH 2.5 acid phosphatase and			
RT	glucose-1-phosphatase."			
RL	J. Bacteriol. 172:5497-5500(1990).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=K12 / MG1655;			
RX	MEDLINE=97426617; PubMed=9278503;			
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,			
RA	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,			
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,			
RA	Mau B., Shao Y.;			
RT	"The complete genome sequence of Escherichia coli K-12."			
RL	Science 277:1453-1474(1997).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=K12;			
RX	MEDLINE=97061202; PubMed=8905232;			
RA	Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,			
RA	Kimura S., Ikeda T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,			
RA	Mori H., Motomura K., Makino K., Masuda S., Miki T., Mizobuchi K.,			
RA	Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,			
RA	Yano M., Horiuchi T.;			
RT	"A 716-kb DNA sequence of the Escherichia coli K-12 genome			
RT	corresponding to the 12.7-28.0 min region on the linkage map."			
RL	DNA Res. 3:137-155(1996).			
RN	[4]			
RP	SEQUENCE OF 1-112 FROM N.A.			
RX	MEDLINE=87271766; PubMed=3038201;			
RA	Touati E., Danchin A.;			
RT	"The structure of the promoter and amino terminal region of the pH			
RT	2.5 acid phosphatase structural gene (appa) of E. coli: a negative			
RT	control of transcription mediated by cyclic AMP."			
RL	Biochimie 69:215-221(1987).			
RN	[5]			



SEQUENCE OF 1-17 FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=92049231; PubMed=1658595;  
RA Dassa J., Fsihi H., Marck C., Dion M., Kieffer-Bontemps M.,  
RA Boquet P.B.;  
RT "A new oxygen-regulated operon in Escherichia coli comprises the  
RT genes for a putative third cytochrome oxidase and for pH 2.5 acid  
RT phosphatase (appa).";  
RL Mol. Gen. Genet. 229:341-352(1991).  
RN [6]  
RP CHARACTERIZATION, AND SEQUENCE OF 23-34.  
RA Greiner R., Jany K.D.;  
RT "Characterization of a phytase from Escherichia coli.";  
RL Biol. Chem. Hoppe-Seyler 372:664-665(1991).  
RN [7]  
RP CHARACTERIZATION, AND SEQUENCE OF 23-35.  
RX MEDLINE=93256556; PubMed=8387749;  
RA Greiner R., Konietzny U., Jany K.D.;  
RT "Purification and characterization of two phytases from Escherichia  
RT coli.";  
RL Arch. Biochem. Biophys. 303:107-113(1993).  
RN [8]  
RP CHARACTERIZATION.  
RX MEDLINE=20161462; PubMed=10696472;  
RA Golovan S., Wang G., Zhang J., Forsberg C.W.;  
RT "Characterization and overproduction of the Escherichia coli appa  
RT encoded bifunctional enzyme that exhibits both phytase and acid  
RT phosphatase activities.";  
RL Can. J. Microbiol. 46:59-71(2000).  
RN [9]  
RP MUTAGENESIS.  
RX MEDLINE=93054596; PubMed=1429631;  
RA Ostanin K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.-M.,  
RA van Etten R.L.;  
RT "Overexpression, site-directed mutagenesis, and mechanism of  
RT Escherichia coli acid phosphatase.";  
RL J. Biol. Chem. 267:22830-22836(1992).  
RN [10]  
RP X-RAY CRYSTALLOGRAPHY (2.05 ANGSTROMS).  
RX MEDLINE=20122624; PubMed=10655611;  
RA Lim D., Golovan S., Forsberg C.W., Jia Z.;  
RT "Crystal structures of Escherichia coli phytase and its complex with  
RT phytate.";  
RL Nat. Struct. Biol. 7:108-113(2000).  
CC -!- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an  
CC alcohol + phosphate.  
CC -!- CATALYTIC ACTIVITY: Myo-inositol hexakisphosphate + H(2)O = 1D-  
CC myo-inositol 1,2,3,4,5-pentakisphosphate + phosphate.  
CC -!- SUBUNIT: Monomer.  
CC -!- SUBCELLULAR LOCATION: Periplasmic.  
CC -!- INDUCTION: IN ADDITION TO CAMP-MEDIATED CONTROL, THIS ENZYME IS  
CC INDUCED WHEN BACTERIAL CULTURES REACH STATIONARY PHASE; ITS  
CC SYNTHESIS IS TRIGGERED BY PHOSPHATE STARVATION OR A SHIFT FROM  
CC AEROBIC TO ANAEROBIC CONDITIONS.  
CC -!- SIMILARITY: Belongs to the histidine acid phosphatase family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL; M58708; AAA72086.1; -;  
DR EMBL; AE000200; AAC74065.1; -;  
DR EMBL; D90735; BAA35745.1; -;  
DR EMBL; X05471; CAA29031.1; -;  
DR EMBL; M58708; -; NOT ANNOTATED\_CDS.  
DR EMBL; S63811; AAB20286.1; -;  
DR PIR; B36733; B36733.  
DR PDB; 1DKL; 03-AUG-00.  
DR PDB; 1DKM; 02-AUG-00.

DR PDB; 1DKN; 03-AUG-00.  
DR PDB; 1DKO; 03-AUG-00.  
DR PDB; 1DKP; 03-AUG-00.  
DR PDB; 1DKQ; 03-AUG-00.  
DR PDB; 1DKR; 03-AUG-00.  
DR PDB; 1DKS; 03-AUG-00.  
DR PDB; 1DKT; 03-AUG-00.  
DR PDB; 1DKU; 03-AUG-00.  
DR PDB; 1DKV; 03-AUG-00.  
DR PDB; 1DKW; 03-AUG-00.  
DR PDB; 1DKX; 03-AUG-00.  
DR PDB; 1DKY; 03-AUG-00.  
DR PDB; 1DKZ; 03-AUG-00.  
DR PDB; 1DK1; 03-AUG-00.  
DR PDB; 1DK2; 03-AUG-00.  
DR PDB; 1DK3; 03-AUG-00.  
DR PDB; 1DK4; 03-AUG-00.  
DR PDB; 1DK5; 03-AUG-00.  
DR PDB; 1DK6; 03-AUG-00.  
DR PDB; 1DK7; 03-AUG-00.  
DR PDB; 1DK8; 03-AUG-00.  
DR PDB; 1DK9; 03-AUG-00.  
DR PDB; 1DKA; 03-AUG-00.  
DR PDB; 1DKB; 03-AUG-00.  
DR PDB; 1DKC; 03-AUG-00.  
DR PDB; 1DKD; 03-AUG-00.  
DR PDB; 1DKE; 03-AUG-00.  
DR PDB; 1DKF; 03-AUG-00.  
DR PDB; 1DKG; 03-AUG-00.  
DR PDB; 1DKH; 03-AUG-00.  
DR PDB; 1DKI; 03-AUG-00.  
DR PDB; 1DKJ; 03-AUG-00.  
DR PDB; 1DKK; 03-AUG-00.  
DR PDB; 1DKL; 03-AUG-00.  
DR PDB; 1DKM; 03-AUG-00.  
DR PDB; 1DKN; 03-AUG-00.  
DR PDB; 1DKO; 03-AUG-00.  
DR PDB; 1DKP; 03-AUG-00.  
DR PDB; 1DKQ; 03-AUG-00.  
DR PDB; 1DKR; 03-AUG-00.  
DR PDB; 1DKS; 03-AUG-00.  
DR PDB; 1DKT; 03-AUG-00.  
DR PDB; 1DKU; 03-AUG-00.  
DR PDB; 1DKV; 03-AUG-00.  
DR PDB; 1DKW; 03-AUG-00.  
DR PDB; 1DKX; 03-AUG-00.  
DR PDB; 1DKY; 03-AUG-00.  
DR PDB; 1DKZ; 03-AUG-00.  
DR PDB; 1DK1; 03-AUG-00.  
DR PDB; 1DK2; 03-AUG-00.  
DR PDB; 1DK3; 03-AUG-00.  
DR PDB; 1DK4; 03-AUG-00.  
DR PDB; 1DK5; 03-AUG-00.  
DR PDB; 1DK6; 03-AUG-00.  
DR PDB; 1DK7; 03-AUG-00.  
DR PDB; 1DK8; 03-AUG-00.  
DR PDB; 1DK9; 03-AUG-00.  
DR PDB; 1DKA; 03-AUG-00.  
DR PDB; 1DKB; 03-AUG-00.  
DR PDB; 1DKC; 03-AUG-00.  
DR PDB; 1DKD; 03-AUG-00.  
DR PDB; 1DKE; 03-AUG-00.  
DR PDB; 1DKF; 03-AUG-00.  
DR PDB; 1DKG; 03-AUG-00.  
DR PDB; 1DKH; 03-AUG-00.  
DR PDB; 1DKI; 03-AUG-00.  
DR PDB; 1DKJ; 03-AUG-00.  
DR PDB; 1DKK; 03-AUG-00.  
DR PDB; 1DKL; 03-AUG-00.  
DR PDB; 1DKM; 03-AUG-00.  
DR PDB; 1DKN; 03-AUG-00.  
DR PDB; 1DKO; 03-AUG-00.  
DR PDB; 1DKP; 03-AUG-00.  
DR PDB; 1DKQ; 03-AUG-00.  
DR PDB; 1DKR; 03-AUG-00.  
DR PDB; 1DKS; 03-AUG-00.  
DR PDB; 1DKT; 03-AUG-00.  
DR PDB; 1DKU; 03-AUG-00.  
DR PDB; 1DKV; 03-AUG-00.  
DR PDB; 1DKW; 03-AUG-00.  
DR PDB; 1DKX; 03-AUG-00.  
DR PDB; 1DKY; 03-AUG-00.  
DR PDB; 1DKZ; 03-AUG-00.  
DR PDB; 1DK1; 03-AUG-00.  
DR PDB; 1DK2; 03-AUG-00.  
DR PDB; 1DK3; 03-AUG-00.  
DR PDB; 1DK4; 03-AUG-00.  
DR PDB; 1DK5; 03-AUG-00.  
DR PDB; 1DK6; 03-AUG-00.  
DR PDB; 1DK7; 03-AUG-00.  
DR PDB; 1DK8; 03-AUG-00.  
DR PDB; 1DK9; 03-AUG-00.  
DR PDB; 1DKA; 03-AUG-00.  
DR PDB; 1DKB; 03-AUG-00.  
DR PDB; 1DKC; 03-AUG-00.  
DR PDB; 1DKD; 03-AUG-00.  
DR PDB; 1DKE; 03-AUG-00.  
DR PDB; 1DKF; 03-AUG-00.  
DR PDB; 1DKG; 03-AUG-00.  
DR PDB; 1DKH; 03-AUG-00.  
DR PDB; 1DKI; 03-AUG-00.  
DR PDB; 1DKJ; 03-AUG-00.  
DR PDB; 1DKK; 03-AUG-00.  
DR PDB; 1DKL; 03-AUG-00.  
DR PDB; 1DKM; 03-AUG-00.  
DR PDB; 1DKN; 03-AUG-00.  
DR PDB; 1DKO; 03-AUG-00.  
DR PDB; 1DKP; 03-AUG-00.  
DR PDB; 1DKQ; 03-AUG-00.  
DR PDB; 1DKR; 03-AUG-00.  
DR PDB; 1DKS; 03-AUG-00.  
DR PDB; 1DKT; 03-AUG-00.  
DR PDB; 1DKU; 03-AUG-00.  
DR PDB; 1DKV; 03-AUG-00.  
DR PDB; 1DKW; 03-AUG-00.  
DR PDB; 1DKX; 03-AUG-00.  
DR PDB; 1DKY; 03-AUG-00.  
DR PDB; 1DKZ; 03-AUG-00.  
DR PDB; 1DK1; 03-AUG-00.  
DR PDB; 1DK2; 03-AUG-00.  
DR PDB; 1DK3; 03-AUG-00.  
DR PDB; 1DK4; 03-AUG-00.  
DR PDB; 1DK5; 03-AUG-00.  
DR PDB; 1DK6; 03-AUG-00.  
DR PDB; 1DK7; 03-AUG-00.  
DR PDB; 1DK8; 03-AUG-00.  
DR PDB; 1DK9; 03-AUG-00.  
DR PDB; 1DKA; 03-AUG-00.  
DR PDB; 1DKB; 03-AUG-00.  
DR PDB; 1DKC; 03-AUG-00.  
DR PDB; 1DKD; 03-AUG-00.  
DR PDB; 1DKE; 03-AUG-00.  
DR PDB; 1DKF; 03-AUG-00.  
DR PDB; 1DKG; 03-AUG-00.  
DR PDB; 1DKH; 03-AUG-00.  
DR PDB; 1DKI; 03-AUG-00.  
DR PDB; 1DKJ; 03-AUG-00.  
DR PDB; 1DKK; 03-AUG-00.  
DR PDB; 1DKL; 03-AUG-00.  
DR PDB; 1DKM; 03-AUG-00.  
DR PDB; 1DKN; 03-AUG-00.  
DR PDB; 1DKO; 03-AUG-00.  
DR PDB; 1DKP; 03-AUG-00.  
DR PDB; 1DKQ; 03-AUG-00.  
DR PDB; 1DKR; 03-AUG-00.  
DR PDB; 1DKS; 03-AUG-00.  
DR PDB; 1DKT; 03-AUG-00.  
DR PDB; 1DKU; 03-AUG-00.  
DR PDB; 1DKV; 03-AUG-00.  
DR PDB; 1DKW; 03-AUG-00.  
DR PDB; 1DKX; 03-AUG-00.  
DR PDB; 1DKY; 03-AUG-00.  
DR PDB; 1DKZ; 03-AUG-00.  
DR PDB; 1DK1; 03-AUG-00.  
DR PDB; 1DK2; 03-AUG-00.  
DR PDB; 1DK3; 03-AUG-00.  
DR PDB; 1DK4; 03-AUG-00.  
DR PDB; 1DK5; 03-AUG-00.  
DR PDB; 1DK6; 03-AUG-00.  
DR PDB; 1DK7; 03-AUG-00.  
DR PDB; 1DK8; 03-AUG-00.  
DR PDB; 1DK9; 03-AUG-00.  
DR PDB; 1DKA; 03-AUG-00.  
DR PDB; 1DKB; 03-AUG-00.  
DR PDB; 1DKC; 03-AUG-00.  
DR PDB; 1DKD; 03-AUG-00.  
DR PDB; 1DKE; 03-AUG-00.  
DR PDB; 1DKF; 03-AUG-00.  
DR PDB; 1DKG; 03-AUG-00.  
DR PDB; 1DKH; 03-AUG-00.  
DR PDB; 1DKI; 03-AUG-00.  
DR PDB; 1DKJ; 03-AUG-00.  
DR PDB; 1DKK; 03-AUG-00.  
DR PDB; 1DKL; 03-AUG-00.  
DR PDB; 1DKM; 03-AUG-00.  
DR PDB; 1DKN; 03-AUG-00.  
DR PDB; 1DKO; 03-AUG-00.  
DR PDB; 1DKP; 03-AUG-00.  
DR PDB; 1DKQ; 03-AUG-00.  
DR PDB; 1DKR; 03-AUG-00.  
DR PDB; 1DKS; 03-AUG-00.  
DR PDB; 1DKT; 03-AUG-00.  
DR PDB; 1DKU; 03-AUG-00.  
DR PDB; 1DKV; 03-AUG-00.  
DR PDB; 1DKW; 03-AUG-00.  
DR PDB; 1DKX; 03-AUG-00.  
DR PDB; 1DKY; 03-AUG-00.  
DR PDB; 1DKZ; 03-AUG-00.  
DR PDB; 1DK1; 03-AUG-00.  
DR PDB; 1DK2; 03-AUG-00.  
DR PDB; 1DK3; 03-AUG-00.  
DR PDB; 1DK4; 03-AUG-00.  
DR PDB; 1DK5; 03-AUG-00.  
DR PDB; 1DK6; 03-AUG-00.  
DR PDB; 1DK7; 03-AUG-00.  
DR PDB; 1DK8; 03-AUG-00.  
DR PDB; 1DK9; 03-AUG-00.  
DR PDB; 1DKA; 03-AUG-00.  
DR PDB; 1DKB; 03-AUG-00.  
DR PDB; 1DKC; 03-AUG-00.  
DR PDB; 1DKD; 03-AUG-00.  
DR PDB; 1DKE; 03-AUG-00.  
DR PDB; 1DKF; 03-AUG-00.  
DR PDB; 1DKG; 03-AUG-00.  
DR PDB; 1DKH; 03-AUG-00.  
DR PDB; 1DKI; 03-AUG-00.  
DR PDB; 1DKJ; 03-AUG-00.  
DR PDB; 1DKK; 03-AUG-00.  
DR PDB; 1DKL; 03-AUG-00.  
DR PDB; 1DKM; 03-AUG-00.  
DR PDB; 1DKN; 03-AUG-00.  
DR PDB; 1DKO; 03-AUG-00.  
DR PDB; 1DKP; 03-AUG-00.  
DR PDB; 1DKQ; 03-AUG-00.  
DR PDB; 1DKR; 03-AUG-00.  
DR PDB; 1DKS; 03-AUG-00.  
DR PDB; 1DKT; 03-AUG-00.  
DR PDB; 1DKU; 03-AUG-00.  
DR PDB; 1DKV; 03-AUG-00.  
DR PDB; 1DKW; 03-AUG-00.  
DR PDB; 1DKX; 03-AUG-00.  
DR PDB; 1DKY; 03-AUG-00.  
DR PDB; 1DKZ; 03-AUG-00.  
DR PDB; 1DK1; 03-AUG-00.  
DR PDB; 1DK2; 03-AUG-00.  
DR PDB; 1DK3; 03-AUG-00.  
DR PDB; 1DK4; 03-AUG-00.  
DR PDB; 1DK5; 03-AUG-00.  
DR PDB; 1DK6; 03-AUG-00.  
DR PDB; 1DK7; 03-AUG-00.  
DR PDB; 1DK8; 03-AUG-00.  
DR PDB; 1DK9; 03-AUG-00.  
DR PDB; 1DKA; 03-AUG-00.  
DR PDB; 1DKB; 03-AUG-00.  
DR PDB; 1DKC; 03-AUG-00.  
DR PDB; 1DKD; 03-AUG-00.  
DR PDB; 1DKE; 03-AUG-00.  
DR PDB; 1DKF; 03-AUG-00.  
DR PDB; 1DKG; 03-AUG-00.  
DR PDB; 1DKH; 03-AUG-00.  
DR PDB; 1DKI; 03-AUG-00.  
DR PDB; 1DKJ; 03-AUG-00.  
DR PDB; 1DKK; 03-AUG-00.  
DR PDB; 1DKL; 03-AUG-00.  
DR PDB; 1DKM; 03-AUG-00.  
DR PDB; 1DKN; 03-AUG-00.  
DR PDB; 1DKO; 03-AUG-00.  
DR PDB; 1DKP; 03-AUG-00.  
DR PDB; 1DKQ; 03-AUG-00.  
DR PDB; 1DKR; 03-AUG-00.  
DR PDB; 1DKS; 03-AUG-00.  
DR PDB; 1DKT; 03-AUG-00.  
DR PDB; 1DKU; 03-AUG-00.  
DR PDB; 1DKV; 03-AUG-00.  
DR PDB; 1DKW; 03-AUG-00.  
DR PDB; 1DKX; 03-AUG-00.  
DR PDB; 1DKY; 03-AUG-00.  
DR PDB; 1DKZ; 03-AUG-00.  
DR PDB; 1DK1; 03-AUG-00.  
DR PDB; 1DK2; 03-AUG-00.  
DR PDB; 1DK3; 03-AUG-00.  
DR PDB; 1DK4; 03-AUG-00.  
DR PDB; 1DK5; 03-AUG-00.  
DR PDB; 1DK6; 03-AUG-00.  
DR PDB; 1DK7; 03-AUG-00.  
DR PDB; 1DK8; 03-AUG-00.  
DR PDB; 1DK9; 03-AUG-00.  
DR PDB; 1DKA; 03-AUG-00.  
DR PDB; 1DKB; 03-AUG-00.  
DR PDB; 1DKC; 03-AUG-00.  
DR PDB; 1DKD; 03-AUG-00.  
DR PDB; 1DKE; 03-AUG-00.  
DR PDB; 1DKF; 03-AUG-00.  
DR PDB; 1DKG; 03-AUG-00.  
DR PDB; 1DKH; 03-AUG-00.  
DR PDB; 1DKI; 03-AUG-00.  
DR PDB; 1DKJ; 03-AUG-00.  
DR PDB; 1DKK; 03-AUG-00.  
DR PDB; 1DKL; 03-AUG-00.  
DR PDB; 1DKM; 03-AUG-00.  
DR PDB; 1DKN; 03-AUG-00.  
DR PDB; 1DKO; 03-AUG-00.  
DR PDB; 1DKP; 03-AUG-00.  
DR PDB; 1DKQ; 03-AUG-00.  
DR PDB; 1DKR; 03-AUG-00.  
DR PDB; 1DKS; 03-AUG-00.  
DR PDB; 1DKT; 03-AUG-00.  
DR PDB; 1DKU; 03-AUG-00.  
DR PDB; 1DKV; 03-AUG-00.  
DR PDB; 1DKW; 03-AUG-00.  
DR PDB; 1DKX; 03-AUG-00.  
DR PDB; 1DKY; 03-AUG-00.  
DR PDB; 1DKZ; 03-AUG-00.  
DR PDB; 1DK1; 03-AUG-00.  
DR PDB; 1DK2; 03-AUG-00.  
DR PDB; 1DK3; 03-AUG-00.  
DR PDB; 1DK4; 03-AUG-00.  
DR PDB; 1DK5; 03-AUG-00.  
DR PDB; 1DK6; 03-AUG-00.  
DR PDB; 1DK7; 03-AUG-00.  
DR PDB; 1DK8; 03-AUG-00.  
DR PDB; 1DK9; 03-AUG-00.  
DR PDB; 1DKA; 03-AUG-00.  
DR PDB; 1DKB; 03-AUG-00.  
DR PDB; 1DKC; 03-AUG-00.  
DR PDB; 1DKD; 03-AUG-00.  
DR PDB; 1DKE; 03-AUG-00.  
DR PDB; 1DKF; 03-AUG-00.  
DR PDB; 1DKG; 03-AUG-00.  
DR PDB; 1DKH; 03-AUG-00.  
DR PDB; 1DKI; 03-AUG-00.  
DR PDB; 1DKJ; 03-AUG-00.  
DR PDB; 1DKK; 03-AUG-00.  
DR PDB; 1DKL; 03-AUG-00.  
DR PDB; 1DKM; 03-AUG-00.  
DR PDB; 1DKN; 03-AUG-00.  
DR PDB; 1DKO; 03-AUG-00.  
DR PDB; 1DKP; 03-AUG-00.  
DR PDB; 1DKQ; 03-AUG-00.  
DR PDB; 1DKR; 03-AUG-00.  
DR PDB; 1DKS; 03-AUG-00.  
DR PDB; 1DKT; 03-AUG-00.  
DR PDB; 1DKU; 03-AUG-00.  
DR PDB; 1DKV; 03-AUG-00.  
DR PDB; 1DKW; 03-AUG-00.  
DR PDB; 1DKX; 03-AUG-00.  
DR PDB; 1DKY; 03-AUG-00.  
DR PDB; 1DKZ; 03-AUG-00.  
DR PDB; 1DK1; 03-AUG-00.  
DR PDB; 1DK2; 03-AUG-00.  
DR PDB; 1DK3; 03-AUG-00.  
DR PDB; 1DK4; 03-AUG-00.  
DR PDB; 1DK5; 03-AUG-00.  
DR PDB; 1DK6; 03-AUG-00.  
DR PDB; 1DK7; 03-AUG-00.  
DR PDB; 1DK8; 03-AUG-00.  
DR PDB; 1DK9; 03-AUG-00.  
DR PDB; 1DKA; 03-AUG-00.  
DR PDB; 1DKB; 03-AUG-00.  
DR PDB; 1DKC; 03-AUG-00.  
DR PDB; 1DKD; 03-AUG-00.  
DR PDB; 1DKE; 03-AUG-00.  
DR PDB; 1DKF; 03-AUG-00.  
DR PDB; 1DKG; 03-AUG-00.  
DR PDB; 1DKH; 03-AUG-00.  
DR PDB; 1DKI; 03-AUG-00.  
DR PDB; 1DKJ; 03-AUG-00.  
DR PDB; 1DKK; 03-AUG-00.  
DR PDB; 1DKL; 03-AUG-00.  
DR PDB; 1DKM; 03-AUG-00.  
DR PDB; 1DKN; 03-AUG-00.  
DR PDB; 1DKO; 03-AUG-00.  
DR PDB; 1DKP; 03-AUG-00.  
DR PDB; 1DKQ; 03-AUG-00.  
DR PDB; 1DKR; 03-AUG-00.  
DR PDB; 1DKS; 03-AUG-00.  
DR PDB; 1DKT; 03-AUG-00.  
DR PDB; 1DKU; 03-AUG-00.  
DR PDB; 1DKV; 03-AUG-00.  
DR PDB; 1DKW; 03-AUG-00.  
DR PDB; 1DKX; 03-AUG-00.  
DR PDB; 1DKY; 03-AUG-00.  
DR PDB; 1DKZ; 03-AUG-00.  
DR PDB; 1DK1; 03-AUG-00.  
DR PDB; 1DK2; 03-AUG-00.  
DR PDB; 1DK3; 03-AUG-00.  
DR PDB; 1DK4; 03-AUG-00.  
DR PDB; 1DK5; 03-AUG-00.  
DR PDB; 1DK6; 03-AUG-00.  
DR PDB; 1DK7; 03-AUG-00.  
DR PDB; 1DK8; 03-AUG-00.  
DR PDB; 1DK9; 03-AUG-00.  
DR PDB; 1DKA; 03-AUG-00.  
DR PDB; 1DKB; 03-AUG-00.  
DR PDB; 1DKC; 03-AUG-00.  
DR PDB; 1DKD; 03-AUG-00.  
DR PDB; 1DKE; 03-AUG-00.  
DR PDB; 1DKF; 03-AUG-00.  
DR PDB; 1DKG; 03-AUG-00.  
DR PDB; 1DKH; 03-AUG-00.  
DR PDB; 1DKI; 03-AUG-00.  
DR PDB; 1DKJ; 03-AUG-00.  
DR PDB; 1DKK; 03-AUG-00.  
DR PDB; 1DKL; 03-AUG-00.  
DR PDB; 1DKM; 03-AUG-00.  
DR PDB; 1DKN; 03-AUG-00.  
DR PDB; 1DKO; 03-AUG-00.  
DR PDB; 1DKP; 03-AUG-00.  
DR PDB; 1DKQ; 03-AUG-00.  
DR PDB; 1DKR; 03-AUG-00.  
DR PDB; 1DKS; 03-AUG-00.  
DR PDB; 1DKT; 03-AUG-00.  
DR PDB; 1DKU; 03-AUG-00.  
DR PDB; 1DKV; 03-AUG-00.  
DR PDB; 1DKW; 03-AUG-00.  
DR PDB; 1DKX; 03-AUG-00.  
DR PDB; 1DKY; 03-AUG-00.  
DR PDB; 1DKZ; 03-AUG-00.  
DR PDB; 1DK1; 03-AUG-00.  
DR PDB; 1DK2; 03-AUG-00.  
DR PDB; 1DK3; 03-AUG-00.  
DR PDB; 1DK4; 03-AUG-00.  
DR PDB; 1DK5; 03-AUG-00.  
DR PDB; 1DK6; 03-AUG-00.  
DR PDB; 1DK7; 03-AUG-00.  
DR PDB; 1DK8; 03-AUG-00.  
DR PDB; 1DK9; 03-AUG-00.  
DR PDB; 1DKA; 03-AUG-00.  
DR PDB; 1DKB; 03-AUG-00.  
DR PDB; 1DKC; 03-AUG-00.  
DR PDB; 1DKD; 03-AUG-00.  
DR PDB; 1DKE; 03-AUG-00.  
DR PDB; 1DKF; 03-AUG-00.  
DR PDB; 1DKG; 03-AUG-00.  
DR PDB; 1DKH; 03-AUG-00.  
DR PDB; 1DKI; 03-AUG-00.  
DR PDB; 1DKJ; 03-AUG-00.  
DR PDB; 1DKK; 03-AUG-00.  
DR PDB; 1DKL; 03-AUG-00.  
DR PDB; 1DKM; 03-AUG-00.  
DR PDB; 1DKN; 03-AUG-00.  
DR PDB; 1DKO; 03-AUG-00.  
DR PDB; 1DKP; 03-AUG-00.  
DR PDB; 1DKQ; 03-AUG-00.  
DR PDB; 1DKR; 03-AUG-00.  
DR PDB; 1DKS; 03-AUG-00.  
DR PDB; 1DKT; 03-AUG-00.  
DR PDB; 1DKU; 03-AUG-00.  
DR PDB; 1DKV; 03-AUG-00.  
DR PDB; 1DKW; 03-AUG-00.  
DR PDB; 1DKX; 03-AUG-00.  
DR PDB; 1DKY; 03-AUG-00.  
DR PDB; 1DKZ; 03-AUG-00.  
DR PDB; 1DK1; 03-AUG-00.  
DR PDB; 1DK2; 03-AUG-00.  
DR PDB; 1DK3; 03-AUG-00.  
DR PDB; 1DK4; 03-AUG-00.  
DR PDB; 1DK5; 03-AUG-00.  
DR PDB; 1DK6; 03-AUG-00.  
DR PDB; 1DK7; 03-AUG-00.  
DR PDB; 1DK8; 03-AUG-00.  
DR PDB; 1DK9; 03-AUG-00.  
DR PDB; 1DKA; 03-AUG-00.  
DR PDB; 1DKB; 03-AUG-00.  
DR PDB; 1DKC; 03-AUG-00.  
DR PDB; 1DKD; 03-AUG-00.  
DR PDB; 1DKE; 03-AUG-00.  
DR PDB; 1DKF; 03-AUG-00.  
DR PDB; 1DKG; 03-AUG-00.  
DR PDB; 1DKH; 03-AUG-00.  
DR PDB; 1DKI; 03-AUG-00.  
DR PDB; 1DKJ; 03-AUG-00.  
DR PDB; 1DKK; 03-AUG-00.  
DR PDB; 1DKL; 03-AUG-00.  
DR PDB; 1DKM; 03-AUG-00.  
DR PDB; 1DKN; 03-AUG-00.  
DR PDB; 1DKO; 03-AUG-00.  
DR PDB; 1DKP; 03-AUG-00.  
DR PDB; 1DKQ; 03-AUG-00.  
DR PDB; 1DKR; 03-AUG-00.  
DR PDB; 1DKS; 03-AUG-00.  
DR PDB; 1DKT; 03-AUG-00.  
DR PDB; 1DKU; 03-AUG-00.  
DR PDB; 1DKV; 03-AUG-00.  
DR PDB; 1DKW; 03-AUG-00.  
DR PDB; 1DKX; 03-AUG-00.  
DR PDB; 1DKY; 03-AUG-00.  
DR PDB; 1DKZ; 03-AUG-00.  
DR PDB; 1DK1; 03-AUG-00.  
DR PDB; 1DK2; 03-AUG-00.  
DR PDB; 1DK3; 03-AUG-00.  
DR PDB; 1DK4; 03-AUG-00.  
DR PDB; 1DK5; 03-AUG-00.  
DR PDB; 1DK6; 03-AUG-00.  
DR PDB; 1DK7; 03-AUG-00.  
DR PDB; 1DK8; 03-AUG-00.  
DR PDB; 1DK9; 03-AUG-00.  
DR PDB; 1DKA; 03-AUG-00.  
DR PDB; 1DKB; 03-AUG-00.  
DR PDB; 1DKC; 03-AUG-00.  
DR PDB; 1DKD; 03-AUG-00.  
DR PDB; 1DKE; 03-AUG-00.  
DR PDB; 1DKF; 03-AUG-00.  
DR PDB; 1DKG; 03-AUG-00.  
DR PDB; 1DKH; 03-AUG-00.  
DR PDB; 1DKI; 03-AUG-00.  
DR PDB; 1DKJ; 03-AUG-00.  
DR PDB; 1DKK; 03-AUG-00.  
DR PDB; 1DKL; 03-AUG-00.  
DR PDB; 1DKM; 03-AUG-00.  
DR PDB; 1DKN; 03-AUG-00.  
DR PDB; 1DKO; 03-AUG-00.  
DR PDB; 1DKP; 03-AUG-00.  
DR PDB; 1DKQ; 03-AUG-00.  
DR PDB; 1DKR; 03-AUG-00.  
DR PDB; 1DKS; 03-AUG-00.  
DR PDB; 1DKT; 03-AUG-00.  
DR PDB; 1DKU; 03-AUG-00.  
DR PDB; 1DKV; 03-AUG-00.  
DR PDB; 1DKW; 03-AUG-00.  
DR PDB; 1DKX; 03-AUG-00.  
DR PDB; 1DKY; 03-AUG-00.  
DR PDB; 1DKZ; 03-AUG-00.  
DR PDB; 1DK1; 03-AUG-00.  
DR PDB; 1DK2; 03-AUG-00.  
DR PDB; 1DK3; 03-AUG-00.  
DR PDB; 1DK4; 03-AUG-00.  
DR PDB; 1DK5; 03-AUG-00.  
DR PDB; 1DK6; 03-AUG-00.  
DR PDB; 1DK7; 03-AUG-00.  
DR PDB; 1DK8; 03-AUG-00.  
DR PDB; 1DK9; 03-AUG-00.  
DR PDB; 1DKA; 03-AUG-00.  
DR PDB; 1DKB; 03-AUG-00.  
DR PDB; 1DKC; 03-AUG-00.  
DR PDB; 1DKD; 03-AUG-00.  
DR PDB; 1DKE; 03-AUG-00.  
DR PDB; 1DKF; 03-AUG-00.  
DR PDB; 1DKG; 03-AUG-00.  
DR PDB; 1DKH; 03-AUG-00.  
DR PDB; 1DKI; 03-AUG-00.  
DR PDB; 1DKJ; 03-AUG-00.  
DR PDB; 1DKK; 03-AUG-00.  
DR PDB; 1DKL; 03-AUG-00.  
DR PDB; 1DKM; 03-AUG-00.  
DR PDB; 1DKN; 03-AUG-00.  
DR PDB; 1DKO; 03-AUG-00.  
DR PDB; 1DKP; 03-AUG-00.  
DR PDB; 1DKQ; 03-AUG-00.  
DR PDB; 1DKR; 03-AUG-00.  
DR PDB; 1DKS; 03-AUG-00.  
DR PDB; 1DKT; 03-AUG-00.  
DR PDB; 1DKU; 03-AUG-00.  
DR PDB; 1DKV; 03-AUG-00.  
DR PDB; 1DKW; 03-AUG-00.  
DR PDB; 1DKX; 03-AUG-00.  
DR PDB; 1DKY; 03-AUG-00.  
DR PDB; 1DKZ; 03-AUG-00.  
DR PDB; 1DK1; 03-AUG-00.  
DR PDB; 1DK2; 03-AUG-00.  
DR PDB; 1DK3; 03-AUG-00.  
DR PDB; 1DK4

```
Query Match      42.7%; Score 939.5; DB 1; Length 432;
Best Local Similarity 47.9%; Pred. No. 1.1e-66;
Matches 196; Conservative 61; Mismatches 137; Indels 15; Gaps 7;

QY 13 LVLMGLAAITAPVA---APSGYTLERVILSHGVRSPTKOTQOLMNDVTPDKWPQ 69
DB 5 LIPFLSLIPLTPQSAFAQSEPE-LKLESVIVSRHGVRAPTKATQOLMNDVTPDAWTP 63
QY 70 VKAGYLTPRGAEVLTMGGFYGYFRSLGLLA-AGCPAEGGVYAQADIDORTLTGOAFL 128
DB 64 VKLGLWTRFGGELIAYLGHYQRLVADGLLAKKGCPSQGVALLADVDERTRKTGEAFA 123
QY 129 DGVAPGCGLTVHNQADLKKTDPLPHVETGVCKLDNAQTDKAIERLGGPDLTVSQRYAK 189
DB 124 AGLAPDCAITVHTQADTSSPDLFNLKTVGCOLDNANVTDAILSRAGGSIADTGHRTQ 183
QY 189 PFAQMGDVLNFAASPYCKSLQOQKTCDFAFHAENNVNKEGTVKLSGSLALSSTLGE 248
DB 184 AFRELERNVLPQSNLCLEKREKQDECSLTQALPSLKVSD--NVSLTGAVSLASMLTE 241
QY 249 IFLQNAQAMPEVAQRUKGAENWVLSLHNAQFNLMKTPYIARHKGTPPLLOQIDTAL 308
DB 242 IFLQNAQAMPEVGRITDQSHQNTLLSLHNAQFYLLQRTPEVARSRAATPLDLIKTAL 301
QY 309 T---LQDAQCKLPISQNRVFLGHDHTNIANLGMIGADWOLPEQDNTPPGGGLVF 365
DB 302 TPHPQKQAYGVTLPTS-----VLFIAGHTNLNLGGLALENLWLPQDNPPEGSELVF 357
QY 366 ELWQNPNDHQRVYAVKMFYQTDQRLNNAEKLDLKNPAGIISVAVAGCE 414
DB 358 ERWRRLSDNSQWIOVSLVFQTLQQRDKTPLSL-NTTPGEVKLTLAGCE 405

RESULT 2
AGP_ECOLI
ID AGP_ECOLI STANDARD; PRT; 413 AA.
AC P19926;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glucose-1-phosphatase precursor (EC 3.1.3.10) (Glpase).
GN AGP OR B1002.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]_TaxID=562;
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-31.
RC STRAIN=K12;
RX MEDLINE=90130318; PubMed=2153660;
RA Pradel E., Marck C., Boquet P.L.;
RT "Nucleotide sequence and transcriptional analysis of the Escherichia
EL coli agp gene encoding periplasmic acid glucose-1-phosphatase.";
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RN Science 277:1453-1474 (1997).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikenoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA
```

```
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155 (1996).
RN [4]
RP SEQUENCE OF 23-34.
RC STRAIN=K12 / ENG2;
RX MEDLINE=97443975; PubMed=9298646;
RA Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RT in the genome of Escherichia coli K-12.";
RL Electrophoresis 18:1259-1313 (1997).
CC -1- FUNCTION: ABSOLUTELY REQUIRED FOR THE GROWTH OF E. COLI IN A HIGH-
CC PHOSPHATE MEDIUM CONTAINING G-1-P AS THE SOLE CARBON SOURCE.
CC -1- CATALYTIC ACTIVITY: D-glucose 1-phosphate + H(2)O = D-glucose +
CC phosphate.
CC -1- ENZYME REGULATION: INDEPENDENT FROM INORGANIC PHOSPHATE
CC AVAILABILITY, AND APPARENTLY SUBMITTED TO CATABOLITE REPRESSION.
CC IT IS POSITIVELY CONTROLLED BY CAMP AND THE CAMP RECEPTOR PROTEIN.
CC -1- SUBUNIT: Homodimer.
CC -1- SUBCELLULAR LOCATION: Periplasmic.
CC -1- MISCELLANEOUS: OPTIMALLY ACTIVE AROUND PH 4.
CC -1- SIMILARITY: Belongs to the histidine acid phosphatase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; M33807; AAA23426.1; --
CC ENBL; AS000202; AAC74087.1; --
CC ENBL; D90737; BAA35769.1; --
CC DR ENBL; D90738; BAA35779.1; --
CC DR PIR; JVO087; JVO087.
CC DR HSSP; P07102; IDKM.
CC DR EcoGene; EGI0033; agp.
CC DR InterPro; IPR000560; HisAc_phsphtase.
CC DR Pfam; PF00328; acid_phosphat; 1.
CC DR PROSITE; PS00616; HIS ACID PHOSPHAT 1; 1.
CC DR PROSITE; PS00778; HIS ACID PHOSPHAT 2; 1.
CC KW Hydrolase; Periplasmic; Signal; Complete proteome.
CC FT SIGNAL 1 22
CC FT CHAIN 23 413 GLUCOSE-1-PHOSPHATASE
CC FT ACT_SITE 40 40 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).
CC FT ACT_SITE 311 311 PROTON DONOR (BY SIMILARITY).
CC SQ SEQUENCE 413 AA; 45683 MW; ADADAD3639D0D6AB CRC64;

Query Match      24.1%; Score 530; DB 1; Length 413;
Best Local Similarity 31.3%; Pred. No. 2.3e-34;
Matches 131; Conservative 82; Mismatches 165; Indels 40; Gaps 11;

QY 10 LSLVLMGLAAITAPVAAPSGYTLERVILSHGVRSPTKOTQOLMNDVTPDKWPQ 68
DB 10 VAGIVLLASNAQAQTV-----PEGYQLQQLVMMRSHNLRAPLANRGSVLEQSTPNKWPEN 64
QY 69 PVKAGYLTPRGAEVLTMGGFYGYFRSLGLLAAG-CPAEGGVYAQADIDORTLTGOAFL 127
DB 65 DVPFGGLTTKGGVLEVMYGHVYRWELAEQGMVSKCEPPYTVYAYANSLQRTVATAQFF 124
QY 128 LDGVAFGCGLTVHNQADLKKTDPLPHVETGVCKLDNAQTDKAIERLGGPDLTVSQRYA 187
DB 125 ITGAFPGCDIPVHOSKMGTMDFTFNPVITDSDSAAFSEQAFAAMEKELS-----KLQUT 178
QY 188 KPPAQMGDVLNFAASPYCKSLQOQKTCDFAFHAENNVNKEGTVKLSGSLALS 243
DB 179 DSYQLLEKLVNRYKDSPPACKKEQCSLVQDQNT-----FSA-----KYQEPGVSGFLKVG 228
QY 244 STLGEIFLQNAQAMP--EVAWQRLKGAENWVLSLHNAQFNLMKTPYIARHKGTPPL 301
DB 244 STLGEIFLQNAQAMP--EVAWQRLKGAENWVLSLHNAQFNLMKTPYIARHKGTPPL 301
```

Db 229 NSLVDAFTLQYEGFPDQVAMGEIKSDQKWVLSKLKNGYQDSLFTSPVARNVAKPLV 288  
QY 302 QQIDTALTQLDAGQKLPISAQNRVFLGGHPTNTANTAGMLG-ADWQLPEQDPTPPG 360  
Db 289 SYIDKALVDR-----TSAPKTVLVGHDSNTASLTALDFPFYQLHQDQERTPIG 339  
QY 361 GGLVFFELQWPNDRHQRVAVKMFYQWQDLNRAEKLDLKNPAGIISVAVAGCNGD 418  
Db 340 GKIVFQRWHSKANRDLMKIEYVYQSAEQLRNADALTQ-APAQRYTLELSGCPIDAD 396

## RESULT 3

ID\_AGP\_SALTY STANDARD; PRT; 413 AA.  
AC O33921;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Glucose-1-phosphatase precursor (EC 3.1.3.10) (GIPase).  
GN AGP OR STM1117.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
OX NCBI\_TaxID=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Lt2 / SGSC1412 / ATCC 700720;  
RX MEDLINE=21534948; PubMed=11677609;  
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
RA Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,  
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nian M.,  
RA Waterston R., Wilson R.K.;  
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
Lt2.";  
RL Nature 413:852-856(2001).  
RN [2]  
RP SEQUENCE OF 1-249 FROM N.A.  
RC STRAIN=TN1379;  
RX MEDLINE=97405892; PubMed=9260936;  
RA Gupta S.D., Wu H.C., Rick P.D.;  
RT "A Salmonella typhimurium genetic locus which confers copper  
RT tolerance on copper-sensitive mutants of Escherichia coli.";  
RL J. Bacteriol. 179:4977-4984(1997).  
CC -1- CATALYTIC ACTIVITY: D-glucose 1-phosphate + H(2)O = D-glucose +  
phosphate.  
CC -1- SUBUNIT: Homodimer (By similarity).  
CC -1- SUBCELLULAR LOCATION: Periplasmic.  
CC -1- SIMILARITY: Belongs to the histidine acid phosphatase family.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; AE008748; AAL20049.1; -  
DR EMBL; U75949; AAC45604.1; -  
DR HSSP; P07102; 1DKL.  
DR StyGene; SGI0595; agp.  
DR InterPro; IPR000560; HisAc\_pheptase.  
DR Pfam; PF00328; acid\_phosphat; 1.  
DR PROSITE; PS00616; HIS\_ACID\_PHOSPHAT\_1; 1.  
DR PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.  
KW Hydrolase; Periplasmic; Signal; Complete proteome.  
FT SIGNAL 1 22  
FT CHAIN 23 413  
FT ACT SITE 40 40  
FT ACT SITE 311 311  
FT PROTON DONOR (BY SIMILARITY).  
FT T -> P (IN REF. 2).  
FT CONFLICT 57 57  
FT CONFLICT 60 63  
FT AWA -> TCPP (IN REF. 2).

FT CONFLICT 74 74 K -> T (IN REF. 2).  
SQ SEQUENCE 413 AA; 45559 MW; E057667CF8A7244C CRC64;  
Query Match 22.7%; Score 499; DB 1; Length 413;  
Best Local Similarity 32.0%; Pred.No.6.4e-32;  
Matches 135; Conservative 77; Mismatches 166; Indels 44; Gaps 14;  
QY 8 VRLGGLVLMLSGLAAITAPVAAEPGSGYTLRVRVILSRHGVRSPTKQTQLMNDVTPDKWP 66  
Db 8 VAVAGAVLLSSAVQAQTT-----PEGYLOQVLMMSRHLRAPLANNNGSVLAQSTPNWP 62  
QY 67 QNPVKAGYLTFRGAELVTLAGGFGYDFYRSLGLLAAG-CPAEGGVYAAQADIDQRTLTGQ 125  
Db 63 ANDVPGGQLTTKGGVLEVTYGHYTRFWLVAQGLIPSGECPAPDPTVYAYANSLQRTVATAQ 122  
QY 126 AFLDGVAPGCGLTVHNOADLKKTDLFHPVETGVCKLDNAQTDKAIBERLGGPLDTVSQR 185  
Db 123 FFIIGAPGCDIPVHQEKMGWMDPTFNPTVITDSSAARQAVQAMEK-----ARSQL 175  
QY 186 YA-KPFAQMGDVLFNFAASPYCKSLQQQKTKCDFAH-----FAANEVNVNKEGKTKVLSGPL 240  
Db 176 HLDESYSKLEQITHYQDPSCKEKHQ-----CSLIDAKDFTFSANYQQ-----EPGVQGPL 225  
QY 241 ALSSTLGRIFLLQNAQAMP--EVAWQRLKGAENWVLSLLHNAQFNLMAKTPYIARHGT 298  
Db 226 KVGNSLVDAFTLQYEGFPDQVAMGGIHDRQKVLKNGYQDSLFTSPVARNVAA 285  
QY 299 FLLOQIDTALTQLDAGQKLPISAQNRVFLGGHPTNTANTAGMLG-ADWQLPEQDPT 357  
Db 286 FLVKYIDKVLAD-----RVSAF-KVTVLVGHDSNIALSLTALDFKPYQLHDQVET 336  
QY 358 PPGGLVFEELWPNDRHQRVAVKMFYQWQDLNRAEKLDLKNPAGIISVAVAGC--EN 415  
Db 337 FIGQGLVFORWHDGNARDLMKIEYVYQSAEQLRNADALTQ-SPAQRVLELKGCPVDA 395  
QY 416 NG 417  
Db 396 NG 397

## RESULT 4

AGP\_PRORE  
ID\_AGP\_PRORE STANDARD; PRT; 417 AA.  
AC Q52309;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Glucose-1-phosphatase precursor (EC 3.1.3.10) (GIPase).  
GN AGP.  
OS Providencia rettgeri.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Providencia.  
OX NCBI\_TaxID=587;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PV7;  
RA Riccio M.L., Chiesurin A., Lombardi G., Satta G.;  
RL Submitted (SEP-1992) to the EMBL/GenBank/DBJ databases.  
CC -1- CATALYTIC ACTIVITY: D-glucose 1-phosphate + H(2)O = D-glucose +  
phosphate.  
CC -1- SUBUNIT: Homodimer (By similarity).  
CC -1- SUBCELLULAR LOCATION: Periplasmic (By similarity).  
CC -1- SIMILARITY: Belongs to the histidine acid phosphatase family.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; X68201; CAA48288.1; -

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U00031; AAK18870.2; --  
 DR HSP; P20646; 1RPA.  
 DR WormPep; B0361.7; CE32100.  
 DR InterPro; IPR000560; HisAc\_phsphtse.  
 DR Pfam; PF00328; acid\_phosphat; 1.  
 DR PROSITE; PS00616; HIS\_ACID\_PHOSPHAT\_1; 1.  
 DR PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.  
 DR PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.  
 KW Hypothetical protein; Signal; Glycoprotein; Hydrolase.  
 FT SIGNAL 1 13  
 FT CHAIN 14 422  
 FT ACT\_SITE 40 40  
 FT ACT\_SITE 80 80  
 FT ACT\_SITE 278 278  
 FT DISULFID 152 363  
 FT DISULFID 205 302  
 FT DISULFID 338 342  
 FT CARBOHYD 104 104  
 FT CARBOHYD 210 210  
 FT CARBOHYD 218 218  
 FT CARBOHYD 312 312  
 FT CARBOHYD 323 323  
 SQ SEQUENCE 422 AA; 48408 MW; 3E29FACB4FAE0CAC CRC64;  
 Query Match 5.7%; Score 125; DB 1; Length 422;  
 Best Local Similarity 20.5%; Pred.No. 0.022;  
 Matches 91; Conservative 58; Mismatches 161; Indels 134; Gaps 20;  
 QY 13 LVLMISGLAATA-----PVAAPSGYTLERVILSRHGVRSPTKQTOLMNDVTPDKWP 66  
 DB 4 LVLVLLGASGINAVVYKEVPIQANTD--TLEYVETVWRHGDRTPAEL-----LFPDDIT 55  
 QY 67 QWPVAGYLTFRGAE-----LVILMGFGYGDYFSRLGLLAAGCPAEGGVYAAQIDQ 118  
 DB 56 KWPEGLGELTEQGAQAQYRLGQWLKRYGSLWGEKN-----RNAIIRSDYN 104  
 QY 119 RTRITGQAFIDGVAP-----GCGILTVHNOADLKKTDPLFHPVETGVCKLDNAQTDKAE 173  
 DB 105 RTLMSAQNAGLFPKYPYIAGGL-----MWQPI--PVHT-----ISKPTDKELYE 148  
 QY 174 RLGGPLDVSQRYAKPQAQMGVNLNFAAPSYCKSLQOQGTCDFAH---FAANEVYNKE 230  
 DB 149 EASCTPAEIMN-----AOW-----KSTKANGIRKKFARELSFFSOKLNPNM 191  
 QY 231 GTKVTLSGLALSSTLGEIFLLQNAQAMPEVAVQRLKAENWVSLSLHNAQFNLMATKP 290  
 DB 192 ELKAT-----WRIEDNLFCEKQNNITWPSMWNSSIFERVDQLYNEVSQLEPHTD 240  
 QY 291 YIARHKGTPLLQIDTALTQLDQAQOKLPISAQNRVILFGGHDNTNANTAGMLGADWQL 350  
 DB 241 TLRLRGTLLEIEIFHRFSK--ASGS---LGKEAKFYASAHDSSTIAALLATLGVFY-- 293  
 QY 351 PEQPDTPPGGLVPFELWQPNHQRVYAVKMFQVTMDQLRNAEKL-----DLKNN 401  
 DB 294 -----DIYPKATCLLI--EMHKLANTRLIRVPHKNETDIDR- 329  
 QY 402 PAGIISVAVAGCENN-----GDD 419  
 DB 330 ---LIEVSIPGCDPCTLQKLGD 350

RESULT 6  
 ID PPAL\_MOUSE  
 AC P24638; O8QZT5;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)

DR PIR; S25627; S25627.  
 DR HSP; S07102; 1DXL.  
 DR InterPro; IPR000560; HisAc\_phsphtse.  
 DR Pfam; PF00328; acid\_phosphat; 1.  
 DR PROSITE; PS00616; HIS\_ACID\_PHOSPHAT\_1; 1.  
 DR PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.  
 KW Hydrolase; Periplasmic; Signal.  
 FT SIGNAL 1 23  
 FT CHAIN 24 417  
 FT ACT\_SITE 42 42  
 FT ACT\_SITE 314 314  
 SQ SEQUENCE 417 AA; 46940 MW; 8CC4CA4F5DBFB90 CRC64;  
 Query Match 21.0%; Score 462; DB 1; Length 417;  
 Best Local Similarity 30.6%; Pred.No. 5.4e-29;  
 Matches 128; Conservative 84; Mismatches 170; Indels 36; Gaps 15;  
 QY 14 VLMLISGLAATAAPVAA-----EPGTYTLERVILSRHGVRSPTKQTOLMNDVTPDKWP 67  
 DB 5 VLTCLLSAALFAPIAPTWASTDNQADMVLDQVLVLSRHLRTPIVNTGILTEVTDKKWPD 64  
 QY 68 WPKVAGYLTFRGAEVTLMGFGYDYSRLGLAAG-CPAEG-GVYAAQIDQDQRTLTGQ 125  
 DB 65 WDAKSGYLTQGGALVYNGHYFREWIDQNKLLADELCTSNEDILYTNLSLQRTIAPA 124  
 QY 126 AFLDGVAPCGGLTVHNOADLKKTDPLFHPVETGVCKLDNAQDKAEIERLGGPLDVSQR 185  
 DB 125 PFAAGAFPOCKVNIHQPEIGKWDVFNPIITNGSPFEKQALAAAMDYLLKG----LSLK 180  
 QY 186 YAKPAQMGDVNLNFAAPSYCKSLQOQGTCDP-AHFAANEVYNKEGTYKVTLSGLALSS 244  
 DB 181 AG--YEELDTVLNIDKQCKT-----DKLCLNDSQKNSFIIEADKE---PGVSGPLKIAN 231  
 QY 245 TLGEITFLQNAQAMP--EYAWORLKAENWVSLSLHNAQFNLMATKPYIARHKGTPLLQ 302  
 DB 232 SAVDAIDLYQEGFPADQVANGLVDPTEKKKLLNTLKNAYQETLFTPKIAKNVAHPILN 291  
 QY 303 QIDTALTQLDQAQOKLPISAQNRVILFGGHDNTNANT-AGMLGADWQLPEQPDTPPGG 361  
 DB 292 YIDKGF-VSD-KGETA-----KFILVGHDSNLSIASMSAMDFKPYLQAOYEHTPIGG 343  
 QY 362 GLVFELWQPNHQRVYAVKMFQVTMDQLRNAEKLDKNNPAGIISVAVAGC--ENNG 417  
 DB 344 KLVFQRTWKQTKQPMKVEYVYQADQLRDNAYLSLETPPKH-VTLEKDCPDKNG 400  
 RESULT 5  
 ID PPAY\_CABEL  
 AC Q10974;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Putative acid phosphatase B0361.7 precursor (EC 3.1.3.2).  
 GN B0361.7.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Paloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1] -----  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Du Z.;  
 RL Submitted (FEB-1994) to the EMBL/GenBank/DBJ databases.  
 RN REVISIONS.  
 RA Waterston R.;  
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
 CC -I- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an  
 CC alcohol + phosphate.  
 CC -I- SIMILARITY: Belongs to the histidine acid phosphatase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

10-OCT-2003 (Rel. 42, Last annotation update)  
Lysosomal acid phosphatase precursor (EC 3.1.3.2) (IAP).  
ACP2.  
Mus musculus (Mouse).  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NCBI\_TaxID=10090;  
[1]  
SEQUENCE FROM N.A.  
TISSUE=Liver;  
MEDLINE=22388257; PubMed=12477932;  
Srausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,  
Raha S.S., Lomellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,  
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
Richards S., Worley K.C., Hsieh S., Garcia A.M., Gay L.J., Hulyk S.W.,  
Vallalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
Fahy J., Heltan E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
"Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences";  
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
[2]  
SEQUENCE OF 3-423 FROM N.A.  
MEDLINE=91282985; PubMed=2059337;  
Geier C., von Figura K., Pohlmann R.;  
"Molecular cloning of the mouse lysosomal acid phosphatase";  
Biol. Chem. Hoppe-Seyler 372:301-304(1991).  
CC -1- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an  
alcohol + phosphate.  
CC -1- SUBCELLULAR LOCATION: Lysosomal.  
CC -1- SIMILARITY: Belongs to the histidine acid phosphatase family.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; BC023343; AAH23343.1; --  
DR EMBL; X57199; CAA40485.1; --  
DR PIR; S14742; S14742.  
DR HSSP; P15309; 2HPA.  
DR MGD; MGI:87882; Acp2.  
DR InterPro; IPR000560; HisAc\_phosphatse.  
DR Pfam; PF00328; acid\_phosphat; 1.  
DR PROSITE; PS00616; HIS\_ACID\_PHOSPHAT\_1; 1.  
DR PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.  
KW Hydrolase; Signal; Glycoprotein; Lysosome.  
FT SIGNAL 1 30  
FT CHAIN 31 423 LYSOSOMAL ACID PHOSPHATASE.  
FT FT 159 370 BY SIMILARITY.  
FT DISULFID 212 310 BY SIMILARITY.  
FT DISULFID 345 349 BY SIMILARITY.  
FT ACT SITE 42 42 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).  
FT ACT SITE 84 84 PROTON DONOR (BY SIMILARITY).  
FT ACT SITE 286 286 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 92 133 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 167 167 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 177 177 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 197 197 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 267 267 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 374 374 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 131 131 N -> S (IN REF. 2).  
SQ SEQUENCE 423 AA; 48508 MW; AFD7A5C90AF2AF4 CRC64;  
Query Match 5.6%; Score 124; DB 1; Length 423;  
Best Local Similarity 22.8%; Pred. No. 0.027;  
Matches 100; Conservative 52; Mismatches 148; Indels 138; Gaps 25;  
QY 11 SGLVLMGLA-AITAPVAEPPSYTLERVVILSRHGVRSPTKTQTLMDVTPDKWPWP 69  
Db 11 AALQFLMGCLTVPFIQAR-----SURFTLLYRHGDRSPVK-TYFKDPYQBEKWFQ-- 63  
QY 70 VKAGYLTPRGA-----EL-VTLMGGFYG----DYFRSLGLLAAGPAGGGVVAQADIDORT 120  
Db 64 -GFQGLTEKGMQLHWELQALRQRYHGLNLSVHRQ-----EYVVRSTDFDRT 110  
QY 121 RLTSQAFLDGVARGCGLTGVNQDLKKTDPLEFHEVETGVCKLDNAQTDKAEERLGGPLD 180  
Db 111 LMSAEANLAGLFP-----PNEVOHFNPNISWQDIP-----VHTVPITEDRL----- 151  
QY 181 TVSQRYAKPFAQMGDVLNFAASPYCKSLQQQKGT-----CDFAHFAANVNV 227  
Db 152 -----LKFPLGCPRYEQQLQNETRQTPYQNRSIQNAQFLNVMVANETGL 195  
QY 228 NKEGTVKTLSPALSLSTLGEIP-----LLQNAQAMPEVAMORLKGAE--NVVSL 275  
Db 196 -----TNVTL-----ETIMNVYDTLCEQTHGLLLPFWASPTV-QLSLQKDFSELFL 243  
QY 276 LSLHNAQFNLMAKTPYIARHKGTEPLLOQIDTALTQLDAQCKLPISAQNRVFLGGHDT 335  
Db 244 FGIH-----EQVQKARLQGVLLAQILKNLTLM--ATTSQFP-----KLLVYSAHDT 288  
QY 336 NIANIAGMLGADWQLPEQDPTPPGGGLVFEELQNPNDHQRVAVKMFYQTMQDLRNAEK 395  
Db 289 TLVALQMALNV-----YNGKQAPYASCHIFELYQEDNGN----FSVEMTF-----RN--- 331  
QY 396 LDLKNNPAGIISVAVAGC 413  
Db 332 -DSKKAPWPLI---LPGC 345  
RESULT 7  
PPAL HUMAN  
ID PPAL HUMAN STANDARD; PRT; 423 AA.  
AC P11117; Q9BTU7;  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Lysosomal acid phosphatase precursor (EC 3.1.3.2) (LAP).  
GN ACP2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RC SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE=Placenta;  
RX MEDLINE=89052645; PubMed=3191910;  
RA Pohlmann R., Krentler C., Schmidt B., Schroeder W., Lorkowski G.,  
Culley J., Mersmann G., Geier C., Waheed A., Gottschalk S.,  
Grzeschik K.H., Hasikik A., von Figura K.;  
"Human lysosomal acid phosphatase: cloning, expression and  
chromosomal assignment";  
EMBO J. 7:2343-2350(1988).  
[2]  
RN SEQUENCE FROM N.A.  
RP TISSUE=Leukocyte;  
RX MEDLINE=89177828; PubMed=2776754;  
RA Geier C., von Figura K., Pohlmann R.;

10-OCT-2003 (Rel. 42, Last annotation update)  
Lysosomal acid phosphatase precursor (EC 3.1.3.2) (IAP).  
ACP2.  
Mus musculus (Mouse).  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NCBI\_TaxID=10090;  
[1]  
SEQUENCE FROM N.A.  
TISSUE=Liver;  
MEDLINE=22388257; PubMed=12477932;  
Srausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,  
Raha S.S., Lomellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,  
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
Richards S., Worley K.C., Hsieh S., Garcia A.M., Gay L.J., Hulyk S.W.,  
Vallalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
Fahy J., Heltan E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
"Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences";  
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
[2]  
SEQUENCE OF 3-423 FROM N.A.  
MEDLINE=91282985; PubMed=2059337;  
Geier C., von Figura K., Pohlmann R.;  
"Molecular cloning of the mouse lysosomal acid phosphatase";  
Biol. Chem. Hoppe-Seyler 372:301-304(1991).  
CC -1- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an  
alcohol + phosphate.  
CC -1- SUBCELLULAR LOCATION: Lysosomal.  
CC -1- SIMILARITY: Belongs to the histidine acid phosphatase family.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; BC023343; AAH23343.1; --  
DR EMBL; X57199; CAA40485.1; --  
DR PIR; S14742; S14742.  
DR HSSP; P15309; 2HPA.  
DR MGD; MGI:87882; Acp2.  
DR InterPro; IPR000560; HisAc\_phosphatse.  
DR Pfam; PF00328; acid\_phosphat; 1.  
DR PROSITE; PS00616; HIS\_ACID\_PHOSPHAT\_1; 1.  
DR PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.  
KW Hydrolase; Signal; Glycoprotein; Lysosome.  
FT SIGNAL 1 30  
FT CHAIN 31 423 LYSOSOMAL ACID PHOSPHATASE.  
FT FT 159 370 BY SIMILARITY.  
FT DISULFID 212 310 BY SIMILARITY.  
FT DISULFID 345 349 BY SIMILARITY.  
FT ACT SITE 42 42 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).  
FT ACT SITE 84 84 PROTON DONOR (BY SIMILARITY).  
FT ACT SITE 286 286 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 92 133 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 167 167 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 177 177 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 197 197 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 267 267 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 374 374 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 131 131 N -> S (IN REF. 2).  
SQ SEQUENCE 423 AA; 48508 MW; AFD7A5C90AF2AF4 CRC64;

"Structure of the human lysosomal acid phosphatase gene."  
 Eur. J. Biochem. 183:611-616(1989).  
 [3]  
 SEQUENCE FROM N.A.  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J.F., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ussid T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loughlano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Boesak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Rulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettner M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grinchow J., Schmutz J., Myers R.M.,  
 RA Butterfield V.S.N., Krzywicki M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RA "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -!- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an  
 CC alcohol + phosphate.  
 CC -!- SUBCELLULAR LOCATION: Lysosomal.  
 CC -!- SIMILARITY: Belongs to the histidine acid phosphatase family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; X12548; CA331064.1; -  
 DR EMBL; X15525; CA333542.1; -  
 DR EMBL; X15526; CA333542.1; JOINED.  
 DR EMBL; X15527; CA333542.1; JOINED.  
 DR EMBL; X15528; CA333542.1; JOINED.  
 DR EMBL; X15529; CA333542.1; JOINED.  
 DR EMBL; X15530; CA333542.1; JOINED.  
 DR EMBL; X15531; CA333542.1; JOINED.  
 DR EMBL; X15532; CA333542.1; JOINED.  
 DR EMBL; X15533; CA333542.1; JOINED.  
 DR EMBL; X15534; CA333542.1; JOINED.  
 DR EMBL; X15535; CA333542.1; JOINED.  
 DR EMBL; X15536; CA333542.1; JOINED.  
 DR EMBL; X15537; CA333542.1; JOINED.  
 DR EMBL; X15538; CA333542.1; JOINED.  
 DR EMBL; X15539; CA333542.1; JOINED.  
 DR EMBL; X15540; CA333542.1; JOINED.  
 DR EMBL; X15541; CA333542.1; JOINED.  
 DR EMBL; X15542; CA333542.1; JOINED.  
 DR EMBL; X15543; CA333542.1; JOINED.  
 DR EMBL; X15544; CA333542.1; JOINED.  
 DR EMBL; X15545; CA333542.1; JOINED.  
 DR EMBL; X15546; CA333542.1; JOINED.  
 DR EMBL; X15547; CA333542.1; JOINED.  
 DR EMBL; X15548; CA333542.1; JOINED.  
 DR EMBL; X15549; CA333542.1; JOINED.  
 DR EMBL; X15550; CA333542.1; JOINED.  
 DR EMBL; X15551; CA333542.1; JOINED.  
 DR EMBL; X15552; CA333542.1; JOINED.  
 DR EMBL; X15553; CA333542.1; JOINED.  
 DR EMBL; X15554; CA333542.1; JOINED.  
 DR EMBL; X15555; CA333542.1; JOINED.  
 DR EMBL; X15556; CA333542.1; JOINED.  
 DR EMBL; X15557; CA333542.1; JOINED.  
 DR EMBL; X15558; CA333542.1; JOINED.  
 DR EMBL; X15559; CA333542.1; JOINED.  
 DR EMBL; X15560; CA333542.1; JOINED.  
 DR EMBL; X15561; CA333542.1; JOINED.  
 DR EMBL; X15562; CA333542.1; JOINED.  
 DR EMBL; X15563; CA333542.1; JOINED.  
 DR EMBL; X15564; CA333542.1; JOINED.  
 DR EMBL; X15565; CA333542.1; JOINED.  
 DR EMBL; X15566; CA333542.1; JOINED.  
 DR EMBL; X15567; CA333542.1; JOINED.  
 DR EMBL; X15568; CA333542.1; JOINED.  
 DR EMBL; X15569; CA333542.1; JOINED.  
 DR EMBL; X15570; CA333542.1; JOINED.  
 DR EMBL; X15571; CA333542.1; JOINED.  
 DR EMBL; X15572; CA333542.1; JOINED.  
 DR EMBL; X15573; CA333542.1; JOINED.  
 DR EMBL; X15574; CA333542.1; JOINED.  
 DR EMBL; X15575; CA333542.1; JOINED.  
 DR EMBL; X15576; CA333542.1; JOINED.  
 DR EMBL; X15577; CA333542.1; JOINED.  
 DR EMBL; X15578; CA333542.1; JOINED.  
 DR EMBL; X15579; CA333542.1; JOINED.  
 DR EMBL; X15580; CA333542.1; JOINED.  
 DR EMBL; X15581; CA333542.1; JOINED.  
 DR EMBL; X15582; CA333542.1; JOINED.  
 DR EMBL; X15583; CA333542.1; JOINED.  
 DR EMBL; X15584; CA333542.1; JOINED.  
 DR EMBL; X15585; CA333542.1; JOINED.  
 DR EMBL; X15586; CA333542.1; JOINED.  
 DR EMBL; X15587; CA333542.1; JOINED.  
 DR EMBL; X15588; CA333542.1; JOINED.  
 DR EMBL; X15589; CA333542.1; JOINED.  
 DR EMBL; X15590; CA333542.1; JOINED.  
 DR EMBL; X15591; CA333542.1; JOINED.  
 DR EMBL; X15592; CA333542.1; JOINED.  
 DR EMBL; X15593; CA333542.1; JOINED.  
 DR EMBL; X15594; CA333542.1; JOINED.  
 DR EMBL; X15595; CA333542.1; JOINED.  
 DR EMBL; X15596; CA333542.1; JOINED.  
 DR EMBL; X15597; CA333542.1; JOINED.  
 DR EMBL; X15598; CA333542.1; JOINED.  
 DR EMBL; X15599; CA333542.1; JOINED.  
 DR EMBL; X15600; CA333542.1; JOINED.  
 DR EMBL; X15601; CA333542.1; JOINED.  
 DR EMBL; X15602; CA333542.1; JOINED.  
 DR EMBL; X15603; CA333542.1; JOINED.  
 DR EMBL; X15604; CA333542.1; JOINED.  
 DR EMBL; X15605; CA333542.1; JOINED.  
 DR EMBL; X15606; CA333542.1; JOINED.  
 DR EMBL; X15607; CA333542.1; JOINED.  
 DR EMBL; X15608; CA333542.1; JOINED.  
 DR EMBL; X15609; CA333542.1; JOINED.  
 DR EMBL; X15610; CA333542.1; JOINED.  
 DR EMBL; X15611; CA333542.1; JOINED.  
 DR EMBL; X15612; CA333542.1; JOINED.  
 DR EMBL; X15613; CA333542.1; JOINED.  
 DR EMBL; X15614; CA333542.1; JOINED.  
 DR EMBL; X15615; CA333542.1; JOINED.  
 DR EMBL; X15616; CA333542.1; JOINED.  
 DR EMBL; X15617; CA333542.1; JOINED.  
 DR EMBL; X15618; CA333542.1; JOINED.  
 DR EMBL; X15619; CA333542.1; JOINED.  
 DR EMBL; X15620; CA333542.1; JOINED.  
 DR EMBL; X15621; CA333542.1; JOINED.  
 DR EMBL; X15622; CA333542.1; JOINED.  
 DR EMBL; X15623; CA333542.1; JOINED.  
 DR EMBL; X15624; CA333542.1; JOINED.  
 DR EMBL; X15625; CA333542.1; JOINED.  
 DR EMBL; X15626; CA333542.1; JOINED.  
 DR EMBL; X15627; CA333542.1; JOINED.  
 DR EMBL; X15628; CA333542.1; JOINED.  
 DR EMBL; X15629; CA333542.1; JOINED.  
 DR EMBL; X15630; CA333542.1; JOINED.  
 DR EMBL; X15631; CA333542.1; JOINED.  
 DR EMBL; X15632; CA333542.1; JOINED.  
 DR EMBL; X15633; CA333542.1; JOINED.  
 DR EMBL; X15634; CA333542.1; JOINED.  
 DR EMBL; X15635; CA333542.1; JOINED.  
 DR EMBL; X15636; CA333542.1; JOINED.  
 DR EMBL; X15637; CA333542.1; JOINED.  
 DR EMBL; X15638; CA333542.1; JOINED.  
 DR EMBL; X15639; CA333542.1; JOINED.  
 DR EMBL; X15640; CA333542.1; JOINED.  
 DR EMBL; X15641; CA333542.1; JOINED.  
 DR EMBL; X15642; CA333542.1; JOINED.  
 DR EMBL; X15643; CA333542.1; JOINED.  
 DR EMBL; X15644; CA333542.1; JOINED.  
 DR EMBL; X15645; CA333542.1; JOINED.  
 DR EMBL; X15646; CA333542.1; JOINED.  
 DR EMBL; X15647; CA333542.1; JOINED.  
 DR EMBL; X15648; CA333542.1; JOINED.  
 DR EMBL; X15649; CA333542.1; JOINED.  
 DR EMBL; X15650; CA333542.1; JOINED.  
 DR EMBL; X15651; CA333542.1; JOINED.  
 DR EMBL; X15652; CA333542.1; JOINED.  
 DR EMBL; X15653; CA333542.1; JOINED.  
 DR EMBL; X15654; CA333542.1; JOINED.  
 DR EMBL; X15655; CA333542.1; JOINED.  
 DR EMBL; X15656; CA333542.1; JOINED.  
 DR EMBL; X15657; CA333542.1; JOINED.  
 DR EMBL; X15658; CA333542.1; JOINED.  
 DR EMBL; X15659; CA333542.1; JOINED.  
 DR EMBL; X15660; CA333542.1; JOINED.  
 DR EMBL; X15661; CA333542.1; JOINED.  
 DR EMBL; X15662; CA333542.1; JOINED.  
 DR EMBL; X15663; CA333542.1; JOINED.  
 DR EMBL; X15664; CA333542.1; JOINED.  
 DR EMBL; X15665; CA333542.1; JOINED.  
 DR EMBL; X15666; CA333542.1; JOINED.  
 DR EMBL; X15667; CA333542.1; JOINED.  
 DR EMBL; X15668; CA333542.1; JOINED.  
 DR EMBL; X15669; CA333542.1; JOINED.  
 DR EMBL; X15670; CA333542.1; JOINED.  
 DR EMBL; X15671; CA333542.1; JOINED.  
 DR EMBL; X15672; CA333542.1; JOINED.  
 DR EMBL; X15673; CA333542.1; JOINED.  
 DR EMBL; X15674; CA333542.1; JOINED.  
 DR EMBL; X15675; CA333542.1; JOINED.  
 DR EMBL; X15676; CA333542.1; JOINED.  
 DR EMBL; X15677; CA333542.1; JOINED.  
 DR EMBL; X15678; CA333542.1; JOINED.  
 DR EMBL; X15679; CA333542.1; JOINED.  
 DR EMBL; X15680; CA333542.1; JOINED.  
 DR EMBL; X15681; CA333542.1; JOINED.  
 DR EMBL; X15682; CA333542.1; JOINED.  
 DR EMBL; X15683; CA333542.1; JOINED.  
 DR EMBL; X15684; CA333542.1; JOINED.  
 DR EMBL; X15685; CA333542.1; JOINED.  
 DR EMBL; X15686; CA333542.1; JOINED.  
 DR EMBL; X15687; CA333542.1; JOINED.  
 DR EMBL; X15688; CA333542.1; JOINED.  
 DR EMBL; X15689; CA333542.1; JOINED.  
 DR EMBL; X15690; CA333542.1; JOINED.  
 DR EMBL; X15691; CA333542.1; JOINED.  
 DR EMBL; X15692; CA333542.1; JOINED.  
 DR EMBL; X15693; CA333542.1; JOINED.  
 DR EMBL; X15694; CA333542.1; JOINED.  
 DR EMBL; X15695; CA333542.1; JOINED.  
 DR EMBL; X15696; CA333542.1; JOINED.  
 DR EMBL; X15697; CA333542.1; JOINED.  
 DR EMBL; X15698; CA333542.1; JOINED.  
 DR EMBL; X15699; CA333542.1; JOINED.  
 DR EMBL; X15700; CA333542.1; JOINED.  
 DR EMBL; X15701; CA333542.1; JOINED.  
 DR EMBL; X15702; CA333542.1; JOINED.  
 DR EMBL; X15703; CA333542.1; JOINED.  
 DR EMBL; X15704; CA333542.1; JOINED.  
 DR EMBL; X15705; CA333542.1; JOINED.  
 DR EMBL; X15706; CA333542.1; JOINED.  
 DR EMBL; X15707; CA333542.1; JOINED.  
 DR EMBL; X15708; CA333542.1; JOINED.  
 DR EMBL; X15709; CA333542.1; JOINED.  
 DR EMBL; X15710; CA333542.1; JOINED.  
 DR EMBL; X15711; CA333542.1; JOINED.  
 DR EMBL; X15712; CA333542.1; JOINED.  
 DR EMBL; X15713; CA333542.1; JOINED.  
 DR EMBL; X15714; CA333542.1; JOINED.  
 DR EMBL; X15715; CA333542.1; JOINED.  
 DR EMBL; X15716; CA333542.1; JOINED.  
 DR EMBL; X15717; CA333542.1; JOINED.  
 DR EMBL; X15718; CA333542.1; JOINED.  
 DR EMBL; X15719; CA333542.1; JOINED.  
 DR EMBL; X15720; CA333542.1; JOINED.  
 DR EMBL; X15721; CA333542.1; JOINED.  
 DR EMBL; X15722; CA333542.1; JOINED.  
 DR EMBL; X15723; CA333542.1; JOINED.  
 DR EMBL; X15724; CA333542.1; JOINED.  
 DR EMBL; X15725; CA333542.1; JOINED.  
 DR EMBL; X15726; CA333542.1; JOINED.  
 DR EMBL; X15727; CA333542.1; JOINED.  
 DR EMBL; X15728; CA333542.1; JOINED.  
 DR EMBL; X15729; CA333542.1; JOINED.  
 DR EMBL; X15730; CA333542.1; JOINED.  
 DR EMBL; X15731; CA333542.1; JOINED.  
 DR EMBL; X15732; CA333542.1; JOINED.  
 DR EMBL; X15733; CA333542.1; JOINED.  
 DR EMBL; X15734; CA333542.1; JOINED.  
 DR EMBL; X15735; CA333542.1; JOINED.  
 DR EMBL; X15736; CA333542.1; JOINED.  
 DR EMBL; X15737; CA333542.1; JOINED.  
 DR EMBL; X15738; CA333542.1; JOINED.  
 DR EMBL; X15739; CA333542.1; JOINED.  
 DR EMBL; X15740; CA333542.1; JOINED.  
 DR EMBL; X15741; CA333542.1; JOINED.  
 DR EMBL; X15742; CA333542.1; JOINED.  
 DR EMBL; X15743; CA333542.1; JOINED.  
 DR EMBL; X15744; CA333542.1; JOINED.  
 DR EMBL; X15745; CA333542.1; JOINED.  
 DR EMBL; X15746; CA333542.1; JOINED.  
 DR EMBL; X15747; CA333542.1; JOINED.  
 DR EMBL; X15748; CA333542.1; JOINED.  
 DR EMBL; X15749; CA333542.1; JOINED.  
 DR EMBL; X15750; CA333542.1; JOINED.  
 DR EMBL; X15751; CA333542.1; JOINED.  
 DR EMBL; X15752; CA333542.1; JOINED.  
 DR EMBL; X15753; CA333542.1; JOINED.  
 DR EMBL; X15754; CA333542.1; JOINED.  
 DR EMBL; X15755; CA333542.1; JOINED.  
 DR EMBL; X15756; CA333542.1; JOINED.  
 DR EMBL; X15757; CA333542.1; JOINED.  
 DR EMBL; X15758; CA333542.1; JOINED.  
 DR EMBL; X15759; CA333542.1; JOINED.  
 DR EMBL; X15760; CA333542.1; JOINED.  
 DR EMBL; X15761; CA333542.1; JOINED.  
 DR EMBL; X15762; CA333542.1; JOINED.  
 DR EMBL; X15763; CA333542.1; JOINED.  
 DR EMBL; X15764; CA333542.1; JOINED.  
 DR EMBL; X15765; CA333542.1; JOINED.  
 DR EMBL; X15766; CA333542.1; JOINED.  
 DR EMBL; X15767; CA333542.1; JOINED.  
 DR EMBL; X15768; CA333542.1; JOINED.  
 DR EMBL; X15769; CA333542.1; JOINED.  
 DR EMBL; X15770; CA333542.1; JOINED.  
 DR EMBL; X15771; CA333542.1; JOINED.  
 DR EMBL; X15772; CA333542.1; JOINED.  
 DR EMBL; X15773; CA333542.1; JOINED.  
 DR EMBL; X15774; CA333542.1; JOINED.  
 DR EMBL; X15775; CA333542.1; JOINED.  
 DR EMBL; X15776; CA333542.1; JOINED.  
 DR EMBL; X15777; CA333542.1; JOINED.  
 DR EMBL; X15778; CA333542.1; JOINED.  
 DR EMBL; X15779; CA333542.1; JOINED.  
 DR EMBL; X15780; CA333542.1; JOINED.  
 DR EMBL; X15781; CA333542.1; JOINED.  
 DR EMBL; X15782; CA333542.1; JOINED.  
 DR EMBL; X15783; CA333542.1; JOINED.  
 DR EMBL; X15784; CA333542.1; JOINED.  
 DR EMBL; X15785; CA333542.1; JOINED.  
 DR EMBL; X15786; CA333542.1; JOINED.  
 DR EMBL; X15787; CA333542.1; JOINED.  
 DR EMBL; X15788; CA333542.1; JOINED.  
 DR EMBL; X15789; CA333542.1; JOINED.  
 DR EMBL; X15790; CA333542.1; JOINED.  
 DR EMBL; X15791; CA333542.1; JOINED.  
 DR EMBL; X15792; CA333542.1; JOINED.  
 DR EMBL; X15793; CA333542.1; JOINED.  
 DR EMBL; X15794; CA333542.1; JOINED.  
 DR EMBL; X15795; CA333542.1; JOINED.  
 DR EMBL; X15796; CA333542.1; JOINED.  
 DR EMBL; X15797; CA333542.1; JOINED.  
 DR EMBL; X15798; CA333542.1; JOINED.  
 DR EMBL; X15799; CA333542.1; JOINED.  
 DR EMBL; X15800; CA333542.1; JOINED.  
 DR EMBL; X15801; CA333542.1; JOINED.  
 DR EMBL; X15802; CA333542.1; JOINED.  
 DR EMBL; X15803; CA333542.1; JOINED.  
 DR EMBL; X15804; CA333542.1; JOINED.  
 DR EMBL; X15805; CA333542.1; JOINED.  
 DR EMBL; X15806; CA333542.1; JOINED.  
 DR EMBL; X15807; CA333542.1; JOINED.  
 DR EMBL; X15808; CA333542.1; JOINED.  
 DR EMBL; X15809; CA333542.1; JOINED.  
 DR EMBL; X15810; CA333542.1; JOINED.  
 DR EMBL; X15811; CA333542.1; JOINED.  
 DR EMBL; X15812; CA333542.1; JOINED.  
 DR EMBL; X15813; CA333542.1; JOINED.  
 DR EMBL; X15814; CA333542.1; JOINED.  
 DR EMBL; X15815; CA333542.1; JOINED.  
 DR EMBL; X15816; CA333542.1; JOINED.  
 DR EMBL; X15817; CA333542.1; JOINED.  
 DR EMBL; X15818; CA333542.1; JOINED.  
 DR EMBL; X15819; CA333542.1; JOINED.  
 DR EMBL; X15820; CA333542.1; JOINED.  
 DR EMBL; X15821; CA333542.1; JOINED.  
 DR EMBL; X15822; CA333542.1; JOINED.  
 DR EMBL; X15823; CA333542.1; JOINED.  
 DR EMBL; X15824; CA333542.1; JOINED.  
 DR EMBL; X15825; CA333542.1; JOINED.  
 DR EMBL; X15826; CA333542.1; JOINED.  
 DR EMBL; X15827; CA333542.1; JOINED.  
 DR EMBL; X15828; CA333542.1; JOINED.  
 DR EMBL; X15829; CA333542.1; JOINED.  
 DR EMBL; X15830; CA333542.1; JOINED.  
 DR EMBL; X15831; CA333542.1; JOINED.  
 DR EMBL; X15832; CA333542.1; JOINED.  
 DR EMBL; X15833; CA333542.1; JOINED.  
 DR EMBL; X15834; CA333542.1; JOINED.  
 DR EMBL; X15835; CA333542.1; JOINED.  
 DR EMBL; X15836; CA333542.1; JOINED.  
 DR EMBL; X15837; CA333542.1; JOINED.  
 DR EMBL; X15838; CA333542.1; JOINED.  
 DR EMBL; X15839; CA333542.1; JOINED.  
 DR EMBL; X15840; CA333542.1; JOINED.  
 DR EMBL; X15841; CA333542.1; JOINED.  
 DR EMBL; X15842; CA333542.1; JOINED.  
 DR EMBL; X15843; CA333542.1; JOINED.  
 DR EMBL; X15844; CA333542.1; JOINED.  
 DR EMBL; X15845; CA333542.1; JOINED.  
 DR EMBL; X15846; CA333542.1; JOINED.  
 DR EMBL; X15847; CA333542.1; JOINED.  
 DR EMBL; X15848; CA333542.1; JOINED.  
 DR EMBL; X15849; CA333542.1; JOINED.  
 DR EMBL; X15850; CA333542.1; JOINED.  
 DR EMBL; X15851; CA333542.1; JOINED.  
 DR EMBL; X15852; CA333542.1; JOINED.  
 DR EMBL; X15853; CA333542.1; JOINED.  
 DR EMBL; X15854; CA333542.1; JOINED.  
 DR EMBL; X15855; CA333542.1; JOINED.  
 DR EMBL; X15856; CA333542.1; JOINED.  
 DR EMBL; X15857; CA333542.1; JOINED.  
 DR EMBL; X15858; CA333542.1; JOINED.  
 DR EMBL; X15859; CA333542.1; JOINED.  
 DR EMBL; X15860; CA333542.1; JOINED.  
 DR EMBL; X15861; CA333542.1; JOINED.  
 DR EMBL; X15862; CA333542.1; JOINED.  
 DR EMBL; X15863; CA333542.1; JOINED.  
 DR EMBL; X15864; CA333542.1; JOINED.  
 DR EMBL; X15865; CA333542.1; JOINED.  
 DR EMBL; X15866; CA333542.1; JOINED.  
 DR EMBL; X15867; CA333542.1; JOINED.  
 DR EMBL; X15868; CA333542.1; JOINED.  
 DR EMBL; X15869; CA333542.1; JOINED.  
 DR EMBL; X15870; CA333542.1; JOINED.  
 DR EMBL; X15871; CA333542.1; JOINED.  
 DR EMBL; X15872; CA333542.1; JOINED.  
 DR EMBL; X15873; CA333542.1; JOINED.  
 DR EMBL; X15874; CA333542.1; JOINED.  
 DR EMBL; X15875; CA333542.1; JOINED.  
 DR EMBL; X15876; CA333542.1; JOINED.  
 DR EMBL; X15877; CA333542.1; JOINED.  
 DR EMBL; X15878; CA333542.1; JOINED.  
 DR EMBL; X15879; CA333542.1; JOINED.  
 DR EMBL; X15880; CA333542.1; JOINED.  
 DR EMBL; X15881; CA333542.1; JOINED.  
 DR EMBL; X15882; CA333542.1; JOINED.  
 DR EMBL; X15883; CA333542.1; JOINED.  
 DR EMBL; X15884; CA333542.1; JOINED.  
 DR EMBL; X15885; CA333542.1; JOINED.  
 DR EMBL; X15886; CA333542.1; JOINED.  
 DR EMBL; X15887; CA333542.1; JOINED.  
 DR EMBL; X15888; CA333542.1; JOINED.  
 DR EMBL; X15889; CA333542.1; JOINED.  
 DR EMBL; X15890; CA333542.1; JOINED.  
 DR EMBL; X15891; CA333542.1; JOINED.  
 DR EMBL; X15892; CA333542.1; JOINED.  
 DR EMBL; X15893; CA333542.1; JOINED.  
 DR EMBL; X15894; CA333542.1; JOINED.  
 DR EMBL; X15895; CA333542.1; JOINED.  
 DR EMBL; X15896; CA333542.1; JOINED.  
 DR EMBL; X15897; CA333542.1; JOINED.  
 DR EMBL; X15898; CA333542.1; JOINED.  
 DR EMBL; X15899; CA333542.1; JOINED.  
 DR EMBL; X15900; CA333542.1; JOINED.  
 DR EMBL; X15901; CA333542.1; JOINED.  
 DR EMBL; X15902; CA333542.1; JOINED.  
 DR EMBL; X15903; CA333542.1; JOINED.  
 DR EMBL; X15904; CA333542.1; JOINED.  
 DR EMBL; X15905; CA333542.1; JOINED.  
 DR EMBL; X15906; CA333542.1; JOINED.  
 DR EMBL; X15907; CA333542.1; JOINED.  
 DR EMBL; X15908; CA333542.1; JOINED.  
 DR EMBL; X15909; CA333542.1; JOINED.  
 DR EMBL; X15910; CA333542.1; JOINED.  
 DR EMBL; X15911; CA333542.1; JOINED.  
 DR EMBL; X15912; CA333542.1; JOINED.  
 DR EMBL; X15913; CA333542.1;



RT of Ty and three trna genes.";  
 RL Yeast 13:171-176 (1997).  
 RN [3]  
 RP SEQUENCE OF 243-541 FROM N.A.  
 RX MEDLINE=97061913; PubMed=8905931;  
 RA Hansen M., Albers M., Backes U., Coblenz A., Leuther H., Neu R.,  
 RA Schreier A., Schaefer B., Zimmermann M., Wolf K.;  
 RT "The sequence of a 23.4 kb segment on the right arm of chromosome VII  
 RT from Saccharomyces cerevisiae reveals CLB6, SPT6, RP28A and NUP57  
 RT genes, a Ty3 element and 11 new open reading frames.";  
 RL Yeast 12:1273-1277 (1996).  
 CC -1- FUNCTION: Part of the nucleoporin complex; required for protein  
 CC transport in the nucleus.  
 CC -1- SUBUNIT: Interacts with NSP1, NUP49 and NUP96.  
 CC -1- SUBCELLULAR LOCATION: Nuclear pore complex.  
 CC -1- DOMAIN: Contains G-L-F-G repeats.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL; X81155; CAA57053.1; -  
 CC EMBL; X83099; CAA58153.1; -  
 CC EMBL; Z72904; CAA97129.1; -  
 CC EMBL; Z72905; CAA97131.1; -  
 CC PIR; S51799; S51799.  
 CC Germline; 141431; -  
 CC SGD; S0003351; NUP57.  
 CC GO; GO:0005643; C:nuclear pore; IDA.  
 CC GO; GO:0006406; P:nuclear pore import; IMP.  
 CC GO; GO:0006606; P:protein-nucleus import; IMP.  
 CC Nuclear protein; Transport; Coiled coil; Repeat.  
 CC FT DOMAIN 76 223 9 X 4 AA REPEATS OF G-L-F-G.  
 CC FT DOMAIN 26 31 POLY-ASN.  
 CC FT DOMAIN 127 130 POLY-THR.  
 CC FT DOMAIN 217 220 POLY-GLY.  
 CC FT DOMAIN 258 266 POLY-GLN.  
 CC FT DOMAIN 277 280 POLY-GLN.  
 CC FT DOMAIN 398 425 COILED COIL (POTENTIAL).  
 CC FT SEQUENCE 541 AA; 57498 MW; B292ADF7B1D7B83C CRC64;  
 SQ  
 Query Match 4.8%; Score 106.5; DB 1; Length 541;  
 Best Local Similarity 23.8%; Pred. No. 0.89;  
 Matches 76; Conservative 41; Mismatches 120; Indels 83; Gaps 18;  
 QY 77 PRGAELVTLMGGVDFYRSLGLAAGCPAEGGVYAQIDQRTLTGQAFLDGVAPGCG 136  
 DB 196 PQGS---TTTGGFLGSGTQNNLTGG---GGLFGSQSQPQ-----TNTAPGLG 237  
 QY 137 LTVENQADLKTDPLPHVETGVCKLDNAQTDKAIEERLGGPLDVTYSQYAKPFAQMGDV 196  
 DB 238 NTVSTQSPFAWSKP-----STG-----SNLQOQO---QQIQVPL-----QQTQALAQOQL 281  
 QY 197 LNFAPSPYCKSLQOQKTC-----DFAHFAANVNVNKGKTVTLGSLALST 245  
 DB 282 SN-----YPOQIQOVLKCKESWDPNNTTKTLRAFYVKNV-----ETEAITYKPK----- 327  
 QY 246 LGEIFLLQNAQAM-----PEVAVORLKGAEWTVSLSLHNAQFNMAKTPYIARH---KG 297  
 DB 328 -GHVLQEWQDAEMKESPTQIPQIYGF-----GLNQNVQVTENVAQRIILNHILEKS 383  
 QY 298 TPLIQ--QIDTALTQLDAQOQKLPISAQNRVFLGQHDNTNIAAGMLG-----ADM 348  
 DB 384 TOLQKHELDATSRILKAQSRNVEI--EKRLKLGTLQATLKNRGLPLGIAEKWMSQF 440  
 QY 349 QLPQPNTPFGGGLVFELW 368  
 DB 441 QTLQRSEDPAGLKTNELW 460

RESULT 9  
 PPAL RAT  
 ID PPAL RAT STANDARD; PRT; 423 AA.  
 AC P20611;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DE 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE Lysosomal acid phosphatase precursor (EC 3.1.3.2) (LAP).  
 GN ACP2.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89350910; PubMed=2764916;  
 RA Himeno M., Fujita H., Noguchi Y., Kono A., Kato K.;  
 RT "Isolation and sequencing of a cDNA clone encoding acid phosphatase  
 RT in rat liver lysosomes.";  
 RL Biochem. Biophys. Res. Commun. 162:1044-1053 (1989).  
 CC -1- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an  
 CC alcohol + phosphate.  
 CC -1- SUBCELLULAR LOCATION: Lysosomal.  
 CC -1- SIMILARITY: Belongs to the histidine acid phosphatase family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL; M27893; AAA0744.1; -  
 CC PIR; A33395; A33395.  
 CC HSRP; P20646; IRPA.  
 CC InterPro; IPR000560; HisAc phosphatase.  
 CC Pfam; PF00328; acid phosphatase; 1.  
 CC PROSITE; PS00616; HIS\_ACID\_PHOSPHAT\_1; 1.  
 CC PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.  
 KW Hydrolase; Signal; Glycoprotein; Lysosome.  
 FT SIGNAL 1 30  
 FT CHAIN 31 423 LYSOSOMAL ACID PHOSPHATASE.  
 FT DISULFID 159 370 BY SIMILARITY.  
 FT DISULFID 212 310 BY SIMILARITY.  
 FT DISULFID 345 349 BY SIMILARITY.  
 FT ACT\_SITE 42 42 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).  
 FT ACT\_SITE 84 84 BY SIMILARITY.  
 FT ACT\_SITE 286 286 PROTON DONOR (BY SIMILARITY).  
 FT CARBOHYD 92 92 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 133 133 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 167 167 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 177 177 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 191 191 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 197 197 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 267 267 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 322 322 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 331 331 N-LINKED (GLCNAC...) (POTENTIAL).  
 SQ SEQUENCE 423 AA; 48319 MW; 6F4C4819CADD4496 CRC64;  
 Query Match 4.8%; Score 106; DB 1; Length 423;  
 Best Local Similarity 23.6%; Pred. No. 0.71;  
 Matches 94; Conservative 49; Mismatches 149; Indels 106; Gaps 23;  
 QY 10 LSGVLMLSGMAATAPVAEPSPGYTLERVVILSRHGVRSPTKQTQLANDVTPDKWQMP 69  
 DB 17 LLGMCLM-----VMPPIQAR-----SLRFVTLVLRHGDGRSPVK-AYPKDPYQBEKWFQ-- 63  
 QY 70 VKAGVLTFRGAELVTLMGGVDFYRSL--GLLAAGCPAEGGVYAQIDQRTLTGQAF 127  
 DB 64 -GFGOLTKEG-----MLQHWELGQALRORYHGFNLNASTHRQ--EVTVRSTDFDRLMSAEAN 117



```
QY 128 LDGAPGGLTVHNOADLKTTDP--LPHVETGVCKLDNAQTDKAIERLGGGLDVTQSQR 185
DB 118 LAGLFP-----PTEVQHFNPNISWQPIP-----VHTVPIEDRL----- 151
QY 186 YAKPFAQGVNLPAAAPYCKSLQOQKT-----CDFAFAANE--VNVNK 229
DB 152 -----LKEFLGCPPEYEQLONETRQTPEYQNNISQNAQFLDMVANETGLMNLTL 200
QY 230 EGTKVTLGGLPALSSTLGEIFLLQNAQAMPV--AWORLKGAEWVSLLSLHNAQFLMA 287
DB 201 E-TWNYVDYTLFCQTHG--LLLPFWASQTVQALSQKDF-SFLFLGHIHQ-----VQ 251
QY 288 KTPYIAHKGPPLQOQIDTALTOLDAGQKLPISAQNRVFLGCHHTNTIANIAGMLGAD 347
DB 252 K-----ARLQGVLLAQILKNTLM--ATTSQFP-----KLVYSAHDTTLVALQMALNV- 299
QY 348 WQLPEQPDNTPGGGLVFLWQNDHQRVYVAKMFYQ 385
DB 300 -----YNGKQAPYASCHIFELVQEDNGN---PSVENYFR 330

RESULT 10
VL1 CRPVK
ID VL1 CRPVK STANDARD; PRT; 505 AA.
AC P03102.
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Major capsid protein L1.
GN L1.
OS Cotton-tail rabbit (shope) papillomavirus (strain Kansas) (CRPV).
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OC NCBI_TaxID=31553;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=85166175; PubMed=2984661;
RA Giri I., Danos O., Yaniv M.;
RL Proc. Natl. Acad. Sci. U.S.A. 82:1580-1584(1985).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K02708; -, NOT ANNOTATED_CDS.
DR PIR; A03643; P1WLRB.
DR InterPro; IPR002210; PV_capsid_L1.
DR InterPro; IPR008975; Viral_cap_coat.
DR Pfam; PF00500; l1; l1.
DR PRINTS; PR00865; HPV_Capsid_L1.
DR ProDom; PD000544; PV_capsid_L1; 1.
KW Coat protein; Late protein.
SQ SEQUENCE 505 AA; 57933 MW; 50FF168DA2F5A38 CRC64;

Query Match
Best Local Similarity 4.7%; Score 103.5; DB 1; Length 505;
Matches 79; Conservative 42; Mismatches 118; Indels 113; Gaps 21;
QY 92 DYRSLG-----LLAAGC-PAEGGVYAA-----DIDQRT-----RLTGOAFLDG--V 131
DB 142 DYRKSMAFPKQVQLLMGLGCVATGSHWAQKCAEDPPQDTCPPIELVNTVIEDGMC 201
QY 132 APCGGLTVN--QADLKKTDPHPVETGVCKLDNAQTDKAIERLGGGLDVTQSQR---Y 186
DB 202 EIGFGANDHKTLQASLSEV-----PLEAQSIKYPDYLMKQKQDGSNMFYARQNY 256
QY 187 AKPFAQM--GDVLNFAASPYCKSLQOQKTCDFAFAANEVNVNKGTVTLGSLALSS 244
```

```
DB 257 ARHFFSRAGGDKENYKSRAYIKRTQMGEA-----NANIATDNYCITPSSGLVSSD 307
QY 245 T--LGEIFLLQNAQAMPE--VAVORLKGAEWVSLLSLHNAQFLMA----- 287
DB 308 SQVFNRYWYLQKQAGWNGVCH-----DNQIFVTVDNTRGTILSLVTKSKEQIKKTHG 361
QY 288 KTPYTA--RHKGTPLLQOQIDTALTOLDAGQKLPISAQNRVFLGCHHTNTIANIAGWL 344
DB 362 KTVHFFSYLRH-----VEEYELQFVLQ-----CKVKLTPENLSYLHSHRPTIIDN----- 407
QY 345 GADWOL-----PEQPDNTPGGGLVFLWQ 369
DB 408 ---WQLSVSAQSPGSLTDQRYLQSIATKCPPEPPKENTIDYKN--YKFEW 454

RESULT 11
6PGD CHLPN
ID 6PGD CHLPN STANDARD; PRT; 479 AA.
AC Q928T3; Q9JQC1;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 6-phosphogluconate dehydrogenase, decarboxylating (EC 1.1.1.44).
GN GND OR CPN0360 OR CP0398 OR CPB0369.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OC NCBI_TaxID=83558;
RN [1]
SEQUENCE FROM N.A.
RX STRAIN=CWL029;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative Genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
RN [2]
SEQUENCE FROM N.A.
RX MEDLINE=99206606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative Genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
RN [3]
SEQUENCE FROM N.A.
RX STRAIN=AR39;
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Winn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
SEQUENCE FROM N.A.
RX STRAIN=J138;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
RN [4]
SEQUENCE FROM N.A.
RX STRAIN=TW-183;
RA Geng M.M., Schumacher A., Muehldorfer I., Bensch K.W., Schaefer K.P.,
RA Schneider S., Pohl T., Essig A., Marre R., Melchers K.;
RT "The genome sequence of Chlamydia pneumoniae TW183 and comparison with
RT other Chlamydia strains based on whole genome sequence analysis.";
RL submitted (May-2003) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: 6-phospho-D-gluconate + NADP(+) = D-ribulose
CC 5-phosphate + CO(2) + NADPH.
CC -1- PATHWAY: Hexose monophosphate shunt.
CC -1- SIMILARITY: Belongs to the 6-phosphogluconate dehydrogenase
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```



Db 547 -----PISAD-VL-----ENLINTAASSFYRETHLATEALNSPPGGFLI 586  
QY 366 ELWQNPD-NHQRVAVKMFYQTMOLNVAEK-----IDLKNNPAGIIS 407  
Db 587 TSDEGKQLVNHPTPLPLFDYSALSCWRELQRLSALLDTPFKARGIIT 635

# RESULT 13

ID DYHC NEUCR STANDARD; PRT; 4367 AA.  
AC P45443;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Dynein heavy chain, cytosolic (DYHC).  
GN RO-1.

OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.  
OX NCBI\_TaxID=5141;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=74-OR23-1A / FGSC 987;  
RX MEDLINE=95014704; PubMed=7929559;  
RA Planam M.; Minke P.F.; Tinsley J.H.; Bruno K.S.;  
RT "Cytosolic dynein and actin-related protein Arp1 are required for  
normal nuclear distribution in filamentous fungi.";  
RL J. Cell Biol. 127:139-149(1994).

CC -!- FUNCTION: Cytosolic dynein acts as a motor for the intracellular  
retrograde motility of vesicles and organelles along microtubules.  
CC Dynein has ATPase activity; the force-producing power stroke is  
thought to occur on release of ADP. Required to maintain uniform  
nuclear distribution in hyphae.

CC -!- SUBUNIT: Consists of at least two heavy chains and a number of  
intermediate and light chains.

CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC (which binds cargo and interacts with other dynein components),  
CC and the head or motor domain. The motor contains six tandemly-

CC linked AAA domains in the head, which form a ring. A stalk-like  
CC structure (formed by two of the coiled coil domains) protrudes  
CC between AAA 4 and AAA 5 and terminates in a microtubule-binding  
CC site. A seventh domain may also contribute to this ring; it is not  
CC clear whether the N-terminus or the C-terminus forms this extra  
CC domain. There are four well-conserved and two non-conserved ATPase  
CC sites, one per AAA domain. Probably only one of these (within AAA  
CC 1) actually hydrolyzes ATP, the others may serve a regulatory  
CC function.

CC -!- SIMILARITY: Belongs to the dynein heavy chain family.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
CC EMBL, L31504; AAA64908.1; --  
DR PIR, B54802; B54802.  
DR InterPro, IPR003593; AAA ATPase.  
DR InterPro, IPR004273; Dynein heavy.  
DR Pfam, PF03028; Dynein heavy; 1.  
DR SMART, SM00382; AAA; 3.

KW Motor protein; Dynein; Microtubule; ATP-binding; Repeat; Coiled coil.  
FT DOMAIN 1 1904 STEM (BY SIMILARITY)  
FT DOMAIN 1905 2130 AAA 1 (BY SIMILARITY).  
FT DOMAIN 2202 2460 AAA 2 (BY SIMILARITY).  
FT DOMAIN 2566 2815 AAA 3 (BY SIMILARITY).  
FT DOMAIN 2909 3179 AAA 4 (BY SIMILARITY).  
FT DOMAIN 3193 3481 STALK (BY SIMILARITY).  
FT DOMAIN 3565 3794 AAA 5 (BY SIMILARITY).

FT DOMAIN 4003 4215 AAA 6 (BY SIMILARITY).  
FT DOMAIN 676 693 COILED COIL (POTENTIAL).  
FT DOMAIN 1176 1215 COILED COIL (POTENTIAL).  
FT DOMAIN 1327 1351 COILED COIL (POTENTIAL).  
FT DOMAIN 1557 1574 COILED COIL (POTENTIAL).  
FT DOMAIN 1637 1668 COILED COIL (POTENTIAL).  
FT DOMAIN 2195 2218 COILED COIL (POTENTIAL).  
FT DOMAIN 3193 3296 COILED COIL (POTENTIAL).  
FT DOMAIN 3423 3481 COILED COIL (POTENTIAL).  
FT DOMAIN 3778 3809 COILED COIL (POTENTIAL).  
FT NP\_BIND 1943 1950 ATP (POTENTIAL).  
FT NP\_BIND 2240 2247 ATP (POTENTIAL).  
FT NP\_BIND 2605 2612 ATP (POTENTIAL).  
FT NP\_BIND 2947 2954 ATP (POTENTIAL).  
SQ SEQUENCE 4367 AA; 495568 MW; 1B10F3E2D170D6F CRC64;

Query Match 4.5%; Score 99.5; DB 1; Length 4367;

Best Local Similarity 18.0%; Pred. No. 49;

Matches 99; Conservative 82; Mismatches 141; Indels 211; Gaps 27;

QY 4 LENRVLGSLVLMISGL-----AATAPVAEPSPGYTLERVVILSRHGVRSSTKQT 54

Db 320 IENQLRSEGVMLTDLILKAKRFOATVSTADTGLKEA---MEKV-----QKYN 365

QY 55 QLANNDVTPDKWPQKVPKAGVLTTPRGAEVLTMGFGYDYFRSLGLLAAGCF----- 105

Db 366 QLMRDPPLDEL-----LSATLT-----KVQESIQIFGLHNLKRLI-----CFYPIRALPL 413

QY 106 --AEGGYAQAIDQRTLTGQAFDGVAPGCGLTVHQAD-----LKTDPFLPHVETGV 159

Db 414 VEAISG-----DLDE-----VLHRLPGTGLVLDYEEFKGVKMQAGSIFRAWDESI 460

QY 160 CKLDNA-----QTDKATEERLGGPLDTSQRYAKPPAQMGDVLPFAFSPYCKSLQQQG 212

Db 461 KEFTNVAEVRTRRNEKFI-----PI-KINPHASLQSKLDYVHFRDN----- 503

QY 213 KTCDFAFHAFANENVNKNKGTQVTLGSLGSLGIFL-----LQNAQAMPEV---AWQRLK- 267

Db 504 -----HEQRTITINVLGPKATVNGIVTASGANGVAVVEIGDVADEVKQAEALDK 557

QY 268 -----GAENVVLSLHN-----AQFILMA 287

Db 558 VLLDCTRETEKRWRAENINERTARVENSIIARLDRLATKANENMFRVSKFNALF 617

QY 288 KTP-----YIARHGTPLLQIDTALTQLDQAQKLPISAQNRVFLFGH----- 333

Db 618 VRPKIRGAIAYQ-TQLIDNVKQAIS-----SLHERFKQYGHSEAHVAQLH 664

QY 334 -----DTNNTANIAMLGADW-----QLPEQPD-----NTPPGG 361

Db 665 DLPPVSGAIIWARQIERQLDQYMKVEQVLGSDWALHTEGQKLQNESDLFRKCLDTRP-- 722

QY 362 GLVFELWQNPDNHQRVAVKMFYQTMOLNVAEKLDLKNPAGIISVAV 410

Db 723 --IFEAWLH-DVQRKQISISGLIFTNIRISA-----GNILELAV 759

## RESULT 14

ID VIE2 HCMVA STANDARD; PRT; 411 AA.  
AC P1983;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 45 kDa immediate-early protein 2 (IE2) (UL122 protein).  
GN UL122.  
OS Human cytomegalovirus (strain AD169).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Betaherpesvirinae; Cytomegalovirus.  
OX NCBI\_TaxID=10360;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90269039; PubMed=2161319;

SEQUENCE FROM N.A.  
 RC STRAIN-Rd / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA McKelvey A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA Klenov J.A., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shiley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 RL Science 269:496-512(1995).  
 CC -!- FUNCTION: SOMEHOW INVOLVED IN THE CYTOCHROME D BRANCH OF AEROBIC  
 CC RESPIRATION. SEEMS TO BE A COMPONENT OF A TRANSPORT SYSTEM  
 CC (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
 CC (Probable).  
 CC -!- SIMILARITY: Belongs to the ABC transporter family. MsbA subfamily.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: U32795; AAC22812.1; -  
 DR PIR: F64186; F64186.  
 DR HSP: P13569; INED.  
 DR TIGR: H11557; -  
 DR InterPro: IPR003593; AAA\_ATPase.  
 DR InterPro: IPR001140; ABC\_TM\_transp.  
 DR InterPro: IPR003439; ABC\_transporter.  
 DR Pfam: PF00664; ABC\_membrane; 1.  
 DR Pfam: PF00005; ABC\_tran; 1.  
 DR SMART: SMO0382; AAA; 1.  
 DR PROSITE: PS50929; ABC\_TM1P; 1.  
 DR PROSITE: PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE: PS00893; ABC\_TRANSPORTER\_2; 1.  
 KW ATP-binding; Transport; Transmembrane; Inner membrane;  
 KW Complete proteome.  
 PT TRANSMEM 28 48 POTENTIAL.  
 FT TRANSMEM 62 82 POTENTIAL.  
 FT TRANSMEM 146 166 POTENTIAL.  
 FT TRANSMEM 167 187 POTENTIAL.  
 FT TRANSMEM 250 270 POTENTIAL.  
 FT TRANSMEM 278 298 POTENTIAL.  
 FT NP\_BIND 383 390 ATP (POTENTIAL).  
 SQ SEQUENCE 586 AA; 65645 MW; A01E8A9920BB5966 CRC64;  
 Query Match 4.5%; Score 98.5; DB 1; Length 586;  
 Best Local Similarity 24.6%; Pred. No. 4.2;  
 Matches 101; Conservative 43; Mismatches 145; Indels 121; Gaps 22;  
 QY 11 SGLVLMGLAATAIPV-----AASPQGYTLERVVILSR-----HGVRSS----- 49  
 DB 165 AGLILM-----ITAPVPLFMIVGIAADNSQKMDTLRLSLAQFLDLRLGLTLRLP 218  
 QY 50 ---PTKQTLMDVTPD--KWPQWPVKAGYLPFGAELVT-----LMGFGYGDYFSLGLL 100  
 DB 219 NRTSEQTEHINATEDFRETTMDVLKLAFLSSAVLEFFTSIALMAYIFG--FSYLGQI 276  
 QY 101 AAGCPAEGVY-----AQADIDQRTLTGQAFLDGVAPCGG----- 136  
 DB 277 -----EFGYNAPLTLTGFFCLILAFEFYQPLDLGTGYHDRAA-GIGADAIVDFLE 329  
 QY 137 ---LTVHQADLKKTDPLFHPVETGVCKLDNAQTDKAIEELRGGLD-----TVS 183  
 DB 330 SDYLTVHQN---EXTISLESASVEISAENLVLSLTGGS---ALTKEPLNFQIPANHNVALVG 383

RA Chee M.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R.,  
 RA Horsnell T., Hutchison C.A. III, Kouratides T., Martignetti J.A.,  
 RA "Freddie E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.;  
 RT "Analysis of the protein-coding content of the sequence of human  
 RL cytomegalovirus strain AD169.";  
 CC Curr. Top. Microbiol. Immunol. 154:125-169(1990).  
 CC -!- FUNCTION: ACTIVATES THE E1F PROMOTER. THIS ACTIVATION IS  
 CC AUGMENTED BY THE E1F PROTEIN. IT DOWN-REGULATES THE TRANSCRIPTION  
 CC OF GENES UNDER THE CONTROL OF THE MAJOR IE PROMOTER.  
 CC -!- SIMILARITY: OF ITS C-TERMINUS TO THE C-TERMINUS OF THE MURINE  
 CC CYTOMEGALOVIRUS IMMEDIATE-EARLY PROTEIN IE3.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: X17403; CAA35324.1; -  
 DR PIR: S09889; EDBEM4.  
 DR InterPro: IPR005028; Herpes IE2\_3.  
 DR Pfam: PF03361; Herpes IE2\_3; 1.  
 KW Early protein; Transcription regulation; Activator; Zinc-finger;  
 KW DNA-binding.  
 FT DOMAIN 89 151 GLY/SER-RICH.  
 FT ZN\_FING 258 284 POTENTIAL.  
 SQ SEQUENCE 411 AA; 44831 MW; 14415DA5AD2C7F31 CRC64;  
 Query Match 4.5%; Score 98.5; DB 1; Length 411;  
 Best Local Similarity 25.3%; Pred. No. 2.7;  
 Matches 78; Conservative 30; Mismatches 119; Indels 81; Gaps 19;  
 QY 87 GGFYGDYFRSLGLLAAGCAEGGYAQAADIDQRTLTGQAFLDGVAPCGGLTVHQAADLK 146  
 DB 128 GGFGG--AASSLLSGHSGHSGGASTGPRKKSKRI-----SELDNE 167  
 QY 147 KTDPLFHPVETGVCKLDNAQTDKAIEELRGPLDTSQRYAK-----PPA--QMG 194  
 DB 168 KVRNIMKDXTPFC--TPNVQTRGRVK-----IDVSRMPRNTNRSLEYKXNLPFTIPSMH 221  
 QY 195 DVLFNFAASPYCKSIQQQCKTCDFAHFAANVNNKGTQVTLGQ--PLALSSLTGIFLIL 252  
 DB 222 QVLEAIIKA-CKTWQVNNKGIIITRNHEVKSVDVAVRCGLTMCNLAISTP----FLM 276  
 QY 253 QNAQAM---PEVAWQRLKGAEN-----WVSLLSLHNAQF-----NLMKATPYIARH 295  
 DB 277 EHTMPVTHPPEVA--QRTADACNEGKAAW--SLKELHQLCPRSSDYRNMI-----H 327  
 QY 296 KGTP--LLQQIDTALTQLDAQAGOKLPISAQNRVLEL--GGHDTNIANIAGMLGADWQLP 351  
 DB 328 AATPVDLLGALNLCPLM-----QKFPQVWVRIFSTNQGFMLPIYETRAKAYAVQGF-- 381  
 QY 352 EQPNTTPP 359  
 DB 382 EQPTETPP 389  
 RESULT 15  
 ID CYDD\_HAEIN STANDARD; PRT; 586 AA.  
 AC P45082;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Transport ATP-binding protein cydd.  
 GN CYDD OR H1157.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Haemophilus.  
 OX NCBI TaxID=727;  
 RN [1]

Qy	184	QRYAKPFAOMGDVLPALSPYCKSLQOQKTCDFAPFAANVNNKGTKVTLSGPLALS	243
Db	384	QSGAGKTSIMNVILGFL--PYEGSLKINGQBLRESNLADWRKHIAWVG-----QNPLLLQ	436
Qy	244	STLGEIFLLQNAQAMPEVAVQRL--KGAENFVSLSLHN-----AQFNLMARKTP	290
Db	437	GTIKENLLGDVQANDEEINQALMESQAKEFTDKLGLHHEIKDGLGISVQQAORLA---	493
Qy	291	YIAR---HKGTPLLOQIDTALTQLDQAQOKLPISAQNRLVFLGHDNTNI	337
Db	494	IARALLRKGDILLLEDEPTA---SLDAQSENVLQALNEA---SQHQTTL	536

Search completed: April 30, 2004, 12:39:27  
Job time : 13.2439 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model  
Run on: April 30, 2004, 12:34:35 ; Search time 38.0488 Seconds  
(without alignments)  
3482.836 Million cell updates/sec

Title: US-10-021-723B-4  
Perfect score: 2200  
Sequence: 1 MSVLENRVRLSGLVLMISGL.....NPAGIISVAVAGCENNGDDK 420

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues  
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL 25.\*
- 1: sp\_archaea.\*
  - 2: sp\_bacteria.\*
  - 3: sp\_fungi.\*
  - 4: sp\_human.\*
  - 5: sp\_invertebrate.\*
  - 6: sp\_mammal.\*
  - 7: sp\_mhc.\*
  - 8: sp\_organalle.\*
  - 9: sp\_phage.\*
  - 10: sp\_plant.\*
  - 11: sp\_rodent.\*
  - 12: sp\_virus.\*
  - 13: sp\_vertebrate.\*
  - 14: sp\_unclassified.\*
  - 15: sp\_virus.\*
  - 16: sp\_bacteriap.\*
  - 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2183	99.2	441	16 Q8ZFP6	Q8zfp6 yersinia pe
2	941.5	42.8	432	2 Q8RKD7	Q8rkd7 escherichia
3	935.5	42.5	432	2 Q8GN88	Q8gn88 escherichia
4	933.5	42.4	432	2 Q8RKD6	Q8rkd6 escherichia
5	933.5	42.4	432	2 Q8RKD8	Q8rkd8 escherichia
6	933.5	42.4	432	2 Q8RXE0	Q8rke0 escherichia
7	932.5	42.4	432	2 Q8RKD9	Q8rkd9 escherichia
8	929.5	42.2	432	2 Q8RKD5	Q8rkd5 escherichia
9	926.5	42.1	434	16 Q8XC29	Q8xc29 escherichia
10	926.5	42.1	446	16 Q8CW75	Q8cw75 escherichia
11	919.5	41.8	432	16 Q7UD08	Q7ud08 shigella fl
12	919.5	41.8	442	16 Q83RW2	Q83rw2 shigella fl
13	532	24.2	413	16 Q8XB26	Q8xb26 escherichia
14	527	24.0	413	16 Q7UD02	Q7ud02 shigella fl
15	527	24.0	421	16 Q83RV6	Q83rv6 shigella fl
16	514.5	23.4	392	16 Q8PP76	Q8pp76 xanthomonas

17	504	22.9	414	16 Q9AAQ4	Q9aaq4 caulobacter
18	500	22.7	443	16 Q8P330	Q8p330 xanthomonas
19	495	22.5	413	16 Q8Z7P1	Q8z7p1 salmoneilla
20	432	19.6	428	2 Q8GD20	Q8gd20 pseudomonas
21	429.5	19.5	421	2 Q84CN9	Q84cn9 klebsiella
22	422	19.2	318	16 Q8CW72	Q8cw72 escherichia
23	421.5	19.2	421	2 Q7WSY1	Q7wsy1 klebsiella
24	413	18.8	435	16 Q8PF53	Q8pf53 xanthomonas
25	332	15.1	414	16 Q8PPX6	Q8ppx6 xanthomonas
26	206.5	9.4	419	2 Q8VQS2	Q8vqs2 klebsiella
27	156	7.1	435	5 Q8MY06	Q8my06 drosophila
28	127.5	5.8	531	5 Q8GNZ3	Q8gnz3 leishmania
29	125.5	5.7	447	5 Q9TW17	Q9tw17 drosophila
30	124.5	5.7	447	5 Q9USU4	Q9usu4 drosophila
31	124.5	5.7	447	5 Q9USU1	Q9usu1 drosophila
32	124.5	5.7	447	5 Q9USV0	Q9usv0 drosophila
33	123.5	5.6	447	5 Q9USU2	Q9usu2 drosophila
34	123.5	5.6	447	5 Q9USU8	Q9usu8 drosophila
35	123.5	5.6	447	5 Q9US58	Q9us58 drosophila
36	123.5	5.6	447	5 Q9US70	Q9us70 drosophila
37	123.5	5.6	447	5 Q9IVM1	Q9ivm1 drosophila
38	123.5	5.6	447	5 Q9US73	Q9us73 drosophila
39	123.5	5.6	447	5 Q9TVX2	Q9tvx2 drosophila
40	123	5.6	447	5 Q9UST5	Q9ust5 drosophila
41	122.5	5.6	447	5 Q9USU5	Q9usu5 drosophila
42	122.5	5.6	447	5 Q97186	Q97186 drosophila
43	122.5	5.6	447	5 Q9USU3	Q9usu3 drosophila
44	122.5	5.6	447	5 Q9TW53	Q9tw53 drosophila
45	122.5	5.6	447	5 Q97187	Q97187 drosophila

ALIGNMENTS

RESULT 1

- Q8ZFP6 PRELIMINARY; PRT; 441 AA.
- AC Q8ZFP6
  - DT 01-MAR-2002 (TrEMBLrel. 20, Created)
  - DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
  - DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
  - DE Probable histidine acid phosphatase (EC 3.1.3.2) (Phosphoanhydride phosphorylase).
  - GN YP01648 OR APPA OR Y1810.
  - OS Yersinia pestis.
  - OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Yersinia.
  - OX NCBI\_TaxID=632;
  - [1]
  - RN SEQUENCE FROM N.A.
  - RP STRAIN=CO-92 / Biovar Orientalis;
  - RC MEDLINE=21470413; PubMed=11586360;
  - RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G., Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L., Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M., Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G., Feilwell T., Hamlin N., Holroyd S., Jagers K., Karlyshev A.V., Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.; "Genome sequence of Yersinia pestis, the causative agent of plague."
  - RL Nature 413:523-527(2001).
  - [2]
  - RN SEQUENCE FROM N.A.
  - RP STRAIN=KIMS / Biovar Mediaevalis;
  - RC MEDLINE=22137863; PubMed=12142430;
  - RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P., Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C., Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V., Straley S.C., McDonough K.A., Niles M.L., Matson J.S., Blattner F.R., Perry R.D.; "Genome sequence of Yersinia pestis KIM."
  - RT J. Bacteriol. 184:4601-4611(2002).
  - RL EMBL; AJ414149; CAC90470.1; -.

DR	EMBL; AE013783; AAM85378.1; -.
DR	PfR; AC0201; AC0201.
DR	GO; GO:000393; F:acid phosphatase activity; IEA.
DR	GO; GO:0016787; F:hydrolase activity; IEA.
DR	InterPro; IPR000560; HisAc phspatse.
DR	Pfam; PF00328; acid phosphat; 1.
DR	PROSITE; PS00616; HIS_Acid PHOSPHAT_1; 1.
KW	Hydrolase; Complete proteome.
SQ	SEQUENCE 441 AA; 47784 MW; CCD15C0A64C98439 CRC64;
<hr/>	
Query Match	99.2%; Score 2183; DB 16; Length 441;
Best Local Similarity	99.3%; Pred. No. 3.8e-169;
Matches 417; Conservative 1; Mismatches 12; Indels 0; Gaps 0;	
QY	1 MSVLNLRVLSGLVLMISGLAAITAPVAAPSQVTLLERVVILSRHGVRSPKTKOTQLMNDV 60
DB	1 MSVLNLRVLSGLVLMISGLAAITAPVAAPSQVTLLERVVILSRHGVRSPKTKOTQLMNDV 60
QY	61 TPDKWQPWPVAGYLTPRGAEVLTMGGFFGYDFRSGLGLLAACCPAEGGVTAQAIDORT 120
DB	61 TPDKWQPWPVAGYLTPRGAEVLTMGGFFGYDFRSGLGLLAACCPAEGGVTAQAIDORT 120
QY	121 RLTSQAFLDGVAPCGGLTVHNADLKKTDPLFHPVETGTGCKLDNAQTDKAIERLGSGPLD 180
DB	121 RLTSQAFLDGVAPCGGLTVHNADLKKTDPLFHPVEAGVKLDAACTDKAIEEQLGGLD 180
QY	191 TVSQRYAKPFAQMGGVNLNFAASPYCKSLQQOGKTCDFAHFAENVNVNKKEGTVLTSGPL 240
DB	191 TVSQRYAKPFAQMGGVNLNFAASPYCKSLQQOGKTCDFAHFAENVNVNKKEGTVLTSGPL 240
QY	241 ALSSTLGEIFLLQNAQAMPVAVQRKGAENWVSLLSHNAQFNLMAKTPTPIARHKGTPL 300
DB	241 ALSSTLGEIFLLQNAQAMPVAVQRKGAENWVSLLSHNAQFNLMAKTPTPIARHKGTPL 300
QY	301 LQQIDTALTOLDRAQGOKLPISAQNRVLFGGHDTNIANTAGMGADWLPEQDPNTPPG 360
DB	301 LQQIDTALTOLDRAQGOKLPISAQNRVLFGGHDTNIANTAGMLGADWLPEQDPNTPPG 360
QY	361 GGLVFELMQNPDNHRVAVKMFTQTDQURNAEKLDKKNPAGIISIVAVACENNNGDDK 420
DB	361 GGLVFELMQNPDNHRVAVKMFTQTDQURNAEKLDKKNPAGIISIVAVACENNNGDDK 420

```

RESULT 2
Q8RK07
ID Q8RKD7 PRELIMINARY; PRT; 432 AA.
AC Q8RKD7;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, last annotation update)
DE Periplasmic phosphoanhydride phosphohydrolase.
OS APPA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90368616; PubMed=2169385;
RA Dassa J., Marcq C., Boquet P.-L.L.;
RT "The complete nucleotide sequence of the Escherichia coli gene appA
RT reveals significant homology between pH 2.5 acid phosphatase and
RT glucose-1-phosphatase.";
RL J. Bacteriol. 172:5497-5500(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93054596; PubMed=1429631;
RA Ostanin K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.M.,
RA Van Etten A.L.;
RT "Overexpression, site-directed mutagenesis, and mechanism of
RT Escherichia coli acid phosphatase.";
RT J. Biol. Chem. 267:22830-22836(1992).
DR EMBL; L03373; AAA00005.1; -.

```

DR	GO; GO:0003993; F:acid phosphatase activity; IEA.
DR	GO; GO:0016787; F:hydrolase activity; IEA.
DR	InterPro; IPR000560; HisAc_phsphtse.
DR	Pfam; PF00328; acid_phosphat_1.
DR	PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR	PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
KW	Hydrolase.
SQL	SEQUENCE 432 AA; 46971 MR; 755D5EAB1AD916A6 CRC64;
Query Match 42.8%; Score 941.5; DB 2; Length 432;	
Best Local Similarity 47.9%; Pred. No. 4.4e-66;	
Matches 196; Conservative 61; Mismatches 137; Indels 15;	
QY	13 LVLMLSGLAAITAPVA--AEPSGYTLERVILSRHGVRSPKTQTQLMNDVTPTR
DB	5 LIPFLSLLLPLTPQSAFAQSEBP-LKLESVIVSRHGVRAPTKATQLKQDVTPT
QY	70 VYAGYLTTPGAEILVTLMGFGYDGFPSLGLLA-AGCPAEGGVYQAADIDORTTR
DB	64 VKLGLWLTTPGGSLIAYLGHYQARLVADGLLAKKCPQSGQVAIIADVDERTR
QY	129 DGVAPGCGITVHNQADLKKTDLPLFHFVETGVCKLDNAQTDKATEERLGGPLDTV
DB	124 AGLAPDCAITVHTQADTSPDPLFNLKTVGCQLDANVTDAILSRAGGSIAAD
QY	189 PFAQMGDVTNFAASPYCKSLQOQKTCDFAPFAANEVNVNKEGTVKLTSLGFLA
DB	184 AFREELRVNLFPSQNLCLKRKODESCSTQALPSELKVSAD--NVSLTGAVSI
QY	249 IFLLQNAQAMPEYAWORLKGAEVNSLLSLHNAQFLMAKTPYIAHKGTPLCLG
DB	242 IFLLQQAQMPPEPGWGRITDSHQWNTLLSLHNAQFVLLQRTPEVARSRAFLPL
QY	309 T---LQLDAQGQKLPISAQNRVILFLGGHDTNINIAAGMAGADWOLPEQPTNP
DB	302 THPPKQAYGVTLPTS----VLFIAGHDTNLNLGALLENWLTLPQGPPTNP
QY	366 ELWQPNDFNRQYVAVQMFQTMWDLQNAEKLDKNNPAGIISVAVAGCE 414
DB	358 ERWRRLSDNSQVTSVLFQTLQMRDKTPLSL-NTPPGVEKVLTLAGCE 405
RESULT 3	
Q8GN88	PRELIMINARY; PRT; 432 AA.
ID	Q8GN88
AC	Q8GN88;
DT	01-MAR-2003 (TrEMBLrel. 23, Created)
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE	APPA.
GN	EScherichia coli.
OS	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriale
OC	Enterobacteriaceae; Escherichia.
OX	NCBI_TaxID=562;
RP	[1]
RP	SEQUENCE FROM N.A.
RA	Zhang Z., He J., Yao B., Zhou Y., Chen Y., Yi Y.
RT	"Production of phytase and acid phosphatase by use of silkworm
RT	bioreactor.";
RL	Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
RL	[2]
RP	SEQUENCE FROM N.A.
RA	Chen Y., Zhu Z., Zhang Z., He J.;
RT	"Cloning and overexpression of phytase gene appA from Escherich
RT	coli.";
RL	Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
RL	EMBL; AF5317219; AN28334.1; --
DR	GO; GO:0003993; F:acid phosphatase activity; IEA.
DR	InterPro; IPR000560; HisAc_phsphtse.
DR	Pfam; PF00328; acid_phosphat_1.
DR	PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR	PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.

RESULT 3	
Q8GN88	PRELIMINARY; PRT; 432 AA.
ID	
AC	O8GN88;
DC	DT 01-WAR-2003 (TrEMBLrel. 23, Created)
DD	DT 01-WAR-2003 (TrEMBLrel. 23, Last sequence update)
DE	DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE	AppA.
OS	Escherichia coli.
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales
OC	Enterobacteriaceae; Escherichia.
OX	NCBI_TaxID=562;
[1]	
RN	SEQUENCE FROM N.A.
RP	Zhang Z., He J., Yao B., Zhou Y., Chen Y., Yi Y.;
RA	"Production of phytase and acid phosphatase by use of silkworm
RT	bioreactor";
RL	Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
RN	[2]
RN	SEQUENCE FROM N.A.
RP	Chen Y., Zhu Z., Zhang Z., He J.;
RA	"Cloning and overexpression of phytase gene appA from Escherichia
RT	coli.";
RL	Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
RN	EMBL: AF537219; AAC29334.1; -
DR	GO: GO:0003993; Fracid phosphatase activity; IEA.
DR	InterPro: IPR000560; HisAc_phsphtse.
DR	Pfam: PF00328; acid_phosphat; 1.
DR	PROSITE: PS00616; HIS ACID PHOSPHAT 1; 1.
DR	PROSITE: PS00778; HIS ACID PHOSPHAT 2; 1.



SQ SEQUENCE 432 AA; 47055 MW; 5B355D76B7377737 CRC64;

Query Match 42.5%; Score 935.5; DB 2; Length 432;  
Best Local Similarity 47.7%; Pred. No. 1.4e-67;  
Matches 195; Conservative 62; Mismatches 137; Indels 15; Gaps 7;

QY 13 LVLMGLAAITAPVA---ABPSGYTLERVVLSRHGVRSPKTKQTLQMDVTPDKWPWP 69  
DB 5 LIPFLSLLIPLTPQSAFAQSEPE-LKLESVVIVSRHGVRAPTKATQLMQDVTPDAMPTWP 63  
QY 70 VKAGVLTFRGAEVLTMGFGYDYSRISGLLA-ACPAEGGVYAAQADIDORTLTGQAF 128  
DB 64 VKLGWLTFRGGELIAYLGHYQRLVADGLLAKGCGSQGVIAIADVDETRTKTGFA 123  
QY 129 DGVAPCGGLTVHQAADLKKTDFLPHFVETGVCKLDNAQTDKAIERLGGPLDVSQRYAK 188  
DB 124 AGLAPDCAITVHTQADTSSPDPLFNPLKTVGCQLDNANVTDAILSRAGGSIAFTGHRQT 183  
QY 189 PFAQMGDVINFAASPYCKSLQOQKTCDFAFHFAENVNKNKGTGKVTLSGPLALSSLTGE 248  
DB 184 AFRELRLVNFQSNLCKRKQDESCSLTQALPSELKVSAD--NVSLTGAVSLASMLTK 241  
QY 249 IFLQNAQAMPEVAWQRKGAENVVSLSLHNAQFNLMAKTFYIARHKGTPLLQOQIDTAL 308  
DB 242 IFLQQAQMGPEFGWGRITDHSQWNTLLSLHNAQFYLLQRTPEVARGATPLDLINKTAL 301  
QY 309 T---LQIDAQOGKLPISAQNRVFLFGGHDNTIANIAGMLGADWQLPQPDNTPPGGLVP 365  
DB 302 THPPPKQAYGVTLPTS----VLFAGHDNTLANLGGALELNTLPGQPDNTPPGGELVF 357  
QY 366 ELWQNDPNHQRVAVKMFYQTMQDLRNAEKLDKNNPAGIISVAVAGCE 414  
DB 358 ERWRRLSDNSQWIVSLVFQTLQQRDKTPLSL-NTPPGEVKLTLAGCE 405

RESULT 5

Q8RKD8 ID Q8RKD8 PRELIMINARY; PRT; 432 AA.

AC Q8RKD8;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Periplasmic phosphoanhydride phosphohydrolase.  
GN APPA.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90368616; PubMed=2168385;  
RA Dassa J., Marck C., Boquet P.-L.L.;  
RT "The complete nucleotide sequence of the Escherichia coli gene appa reveals significant homology between pH 2.5 acid phosphatase and glucose-1-phosphatase."  
RT J. Bacteriol. 172:5497-5500(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93054596; PubMed=1429631;  
RA Ostalan K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.M.,  
RA Van Etten R.L.;  
RT "Overexpression, site-directed mutagenesis, and mechanism of Escherichia coli acid phosphatase."  
RT J. Biol. Chem. 267:22830-22836(1992).  
RL EMBL; L03372; AAA00004.1;  
DR GO; GO:0003993; F:acid phosphatase activity; IEA.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR000560; HisAc\_phspatse.  
DR Pfam; PF00328; acid\_phosphat; 1.  
DR PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.  
KW Hydrolase.  
SQ SEQUENCE 432 AA; 46971 MW; 9A85536B57FCCFB5 CRC64;

Query Match 42.4%; Score 933.5; DB 2; Length 432;

SQ SEQUENCE 432 AA; 47055 MW; 5B355D76B7377737 CRC64;

Query Match 42.5%; Score 935.5; DB 2; Length 432;  
Best Local Similarity 47.7%; Pred. No. 1.4e-67;  
Matches 195; Conservative 62; Mismatches 137; Indels 15; Gaps 7;

QY 13 LVLMGLAAITAPVA---ABPSGYTLERVVLSRHGVRSPKTKQTLQMDVTPDKWPWP 69  
DB 5 LIPFLSLLIPLTPQSAFAQSEPE-LKLESVVIVSRHGVRAPTKATQLMQDVTPDAMPTWP 63  
QY 70 VKAGVLTFRGAEVLTMGFGYDYSRISGLLA-ACPAEGGVYAAQADIDORTLTGQAF 128  
DB 64 VKLGWLTFRGGELIAYLGHYQRLVADGLLAKGCGSQGVIAIADVDETRTKTGFA 123  
QY 129 DGVAPCGGLTVHQAADLKKTDFLPHFVETGVCKLDNAQTDKAIERLGGPLDVSQRYAK 188  
DB 124 AGLAPDCAITVHTQADTSSPDPLFNPLKTVGCQLDNANVTDAILSRAGGSIAFTGHRQT 183  
QY 189 PFAQMGDVINFAASPYCKSLQOQKTCDFAFHFAENVNKNKGTGKVTLSGPLALSSLTGE 248  
DB 184 AFRELRLVNFQSNLCKRKQDESCSLTQALPSELKVSAD--NVSLTGAVSLASMLTK 241  
QY 249 IFLQNAQAMPEVAWQRKGAENVVSLSLHNAQFNLMAKTFYIARHKGTPLLQOQIDTAL 308  
DB 242 IFLQQAQMGPEFGWGRITDHSQWNTLLSLHNAQFYLLQRTPEVARGATPLDLINKTAL 301  
QY 309 T---LQIDAQOGKLPISAQNRVFLFGGHDNTIANIAGMLGADWQLPQPDNTPPGGLVP 365  
DB 302 THPPPKQAYGVTLPTS----VLFAGHDNTLANLGGALELNTLPGQPDNTPPGGELVF 357  
QY 366 ELWQNDPNHQRVAVKMFYQTMQDLRNAEKLDKNNPAGIISVAVAGCE 414  
DB 358 ERWRRLSDNSQWIVSLVFQTLQQRDKTPLSL-NTPPGEVKLTLAGCE 405

RESULT 4

Q8RKD6 ID Q8RKD6 PRELIMINARY; PRT; 432 AA.

AC Q8RKD6;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Periplasmic phosphoanhydride phosphohydrolase.  
GN APPA.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90368616; PubMed=2168385;  
RA Dassa J., Marck C., Boquet P.-L.L.;  
RT "The complete nucleotide sequence of the Escherichia coli gene appa reveals significant homology between pH 2.5 acid phosphatase and glucose-1-phosphatase."  
RT J. Bacteriol. 172:5497-5500(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93054596; PubMed=1429631;  
RA Ostalan K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.M.,  
RA Van Etten R.L.;  
RT "Overexpression, site-directed mutagenesis, and mechanism of Escherichia coli acid phosphatase."  
RT J. Biol. Chem. 267:22830-22836(1992).  
RL EMBL; L03374; AAA00006.1;  
DR GO; GO:0003993; F:acid phosphatase activity; IEA.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR000560; HisAc\_phspatse.  
DR Pfam; PF00328; acid\_phosphat; 1.  
DR PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.  
KW Hydrolase.  
SQ SEQUENCE 432 AA; 46971 MW; AF86C41EA6193AC5 CRC64;

Best Local Similarity 47.7%; Pred. No. 2e-67;  
Matches 195; Conservative 61; Mismatches 138; Indels 15; Gaps 7;

QY 13 LVLMISGLAIAITAPVA---AEPGTYTLERVVILSRHGVSPKTKQTOLMNDVTPDKWPQP 69  
Db 5 LIPFLSLLIPLTPQSAFAQSEPE-LKLESVVIVSAHGVRAPTKATQLMQDVTDPDWTWP 63  
QY 70 VKAGYTPRGAELVTMGPGYGYDFRSLGLLA-AGCPAEGVYAQADIDORTLTGQAF 128  
Db 64 VKLGWLTFRGGELIAYLGHYQORLQVADGLLAKKGCPSQGVAIADVDERTKRTGEAFA 123  
QY 129 DGVAAGGLTVHNOADLKKTDPLFHPVETGVCKLDNAQTDKAEERLGGPLDTVVSQRYAK 188  
Db 124 AGLAPDCAITVHTQADTSSPDPLFNPLKTVGCOLDNANVTDAILSRAGGSIAFTGHRQT 183  
QY 189 PFAQMGDVLPFAASPYCKSLQOQKTCDFAHFAANVNNKGTQVTLSPGLASLTGE 248  
Db 184 AFRELRLVLPQSNLCKREKQDECSLTQALPSELKVSAD--NVSLTAGVSLASMLTE 241  
QY 249 IFLQNAQAMPEVANOELKGAENWVSLSLHNAQFNLMAKTPYIARHKGTPLLQOQIDTAL 308  
Db 242 IFLQQAQMGPEPGWGRITDQSHQNTLLSLHNAQFYLLQRTPEVARSATPLDLIKTAL 301  
QY 309 T---LQIDAQOQKLPISAQNRVFLFGGHDNTNANIAGMLGADWQLPEQPDNTPPGGGLVF 365  
Db 302 TPHPPOKQAYGVTLPTS---VLFIAGHDTNLANLGGALELNLWTLPGQPDNTPPGGGLVF 357  
QY 366 ELWQNPNDHORYVAVKMFYQTMOLRVAEKLDLKNPAGIISVAVAGCE 414  
Db 358 ERWRRLSDNSQWIOVSLVFQTLQOQMRDKTPLSL-NTFPGGEVKLTAGCE 405

RESULT 6  
Q8RKEO PRELIMINARY; PRT; 432 AA.

ID Q8RKEO AC Q8RKEO  
DT 01-JUN-2002 (TREMBlrel. 21, Created)  
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
DE Periplasmic phosphoanhydride phosphohydrolase.  
GN APPA.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90368616; PubMed=2168385;  
RA Dassa J., Marck C., Boquet P.-L.L.;  
RT "The complete nucleotide sequence of the Escherichia coli gene appA  
RT reveals significant homology between pH 2.5 acid phosphatase and  
RT Glucose-1-phosphatase.";  
RL J. Bacteriol. 172:5497-5500(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93054596; PubMed=1429631;  
RA Ostanin K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.M.,  
RA Van Etten R.L.;  
RT "Overexpression, site-directed mutagenesis, and mechanism of  
RT Escherichia coli acid phosphatase.";  
RL J. Biol. Chem. 267:22830-22836(1992).  
DR EMBL; L03370; AAA00002.1;  
DR GO; GO:0003993; F:acid phosphatase activity; IEA.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR000560; HisAc\_phsphtse.  
DR Pfam; PF00328; acid\_phosphat; 1.  
DR PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.  
KW Hydrolase.  
SQ SEQUENCE 432 AA; 46971 MW; 5BBD632D4682EABF CRC64;

Query Match 42.4%; Score 933.5; DB 2; Length 432;  
Best Local Similarity 47.7%; Pred. No. 2e-67;  
Matches 195; Conservative 61; Mismatches 138; Indels 15; Gaps 7;

Best Local Similarity 47.7%; Pred. No. 2e-67;  
Matches 195; Conservative 61; Mismatches 138; Indels 15; Gaps 7;

QY 13 LVLMISGLAIAITAPVA---AEPGTYTLERVVILSRHGVSPKTKQTOLMNDVTPDKWPQP 69  
Db 5 LIPFLSLLIPLTPQSAFAQSEPE-LKLESVVIVSAHGVRAPTKATQLMQDVTDPDWTWP 63  
QY 70 VKAGYTPRGAELVTMGPGYGYDFRSLGLLA-AGCPAEGVYAQADIDORTLTGQAF 128  
Db 64 VKLGWLTFRGGELIAYLGHYQORLQVADGLLAKKGCPSQGVAIADVDERTKRTGEAFA 123  
QY 129 DGVAAGGLTVHNOADLKKTDPLFHPVETGVCKLDNAQTDKAEERLGGPLDTVVSQRYAK 188  
Db 124 AGLAPDCAITVHTQADTSSPDPLFNPLKTVGCOLDNANVTDAILSRAGGSIAFTGHRQT 183  
QY 189 PFAQMGDVLPFAASPYCKSLQOQKTCDFAHFAANVNNKGTQVTLSPGLASLTGE 248  
Db 184 AFRELRLVLPQSNLCKREKQDECSLTQALPSELKVSAD--NVSLTAGVSLASMLTE 241  
QY 249 IFLQNAQAMPEVANOELKGAENWVSLSLHNAQFNLMAKTPYIARHKGTPLLQOQIDTAL 308  
Db 242 IFLQQAQMGPEPGWGRITDQSHQNTLLSLHNAQFYLLQRTPEVARSATPLDLIKTAL 301  
QY 309 T---LQIDAQOQKLPISAQNRVFLFGGHDNTNANIAGMLGADWQLPEQPDNTPPGGGLVF 365  
Db 302 TPHPPOKQAYGVTLPTS---VLFIAGHDTNLANLGGALELNLWTLPGQPDNTPPGGGLVF 357  
QY 366 ELWQNPNDHORYVAVKMFYQTMOLRVAEKLDLKNPAGIISVAVAGCE 414  
Db 358 ERWRRLSDNSQWIOVSLVFQTLQOQMRDKTPLSL-NTFPGGEVKLTAGCE 405

RESULT 7  
Q8RKEO PRELIMINARY; PRT; 432 AA.

ID Q8RKEO AC Q8RKEO  
DT 01-JUN-2002 (TREMBlrel. 21, Created)  
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
DE Periplasmic phosphoanhydride phosphohydrolase.  
GN APPA.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90368616; PubMed=2168385;  
RA Dassa J., Marck C., Boquet P.-L.L.;  
RT "The complete nucleotide sequence of the Escherichia coli gene appA  
RT reveals significant homology between pH 2.5 acid phosphatase and  
RT Glucose-1-phosphatase.";  
RL J. Bacteriol. 172:5497-5500(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93054596; PubMed=1429631;  
RA Ostanin K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.M.,  
RA Van Etten R.L.;  
RT "Overexpression, site-directed mutagenesis, and mechanism of  
RT Escherichia coli acid phosphatase.";  
RL J. Biol. Chem. 267:22830-22836(1992).  
DR EMBL; L03370; AAA00003.1;  
DR GO; GO:0003993; F:acid phosphatase activity; IEA.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR000560; HisAc\_phsphtse.  
DR Pfam; PF00328; acid\_phosphat; 1.  
DR PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.  
KW Hydrolase.  
SQ SEQUENCE 432 AA; 47033 MW; 9F29B9DF9C368175 CRC64;

Query Match 42.4%; Score 932.5; DB 2; Length 432;  
Best Local Similarity 47.7%; Pred. No. 2.4e-67;  
Matches 195; Conservative 62; Mismatches 137; Indels 15; Gaps 7;

QY 13 LVLMISGLAIAITAPVA---AEPGTYTLERVVILSRHGVSPKTKQTOLMNDVTPDKWPQP 69

Db 5 LIPFLSLPLTPQSAFAQSEPE-LKLESVIVSRNGVRAPTKATQMLQMDVTPDAWPTWP 63  
 QY 70 VKAGVLTTPRGAEVLTLMGGFYGDYFRSLGLLA-AGCPAEGGVYAQADIDQRTLTGQAF 128  
 Db 64 VKLGLWLTTPRGELIAYLGHYQRLVADGLLAKKCPQSGQVAILADVDERTRKTGEAFA 123  
 QY 129 DGVAFCGLTVHNOADLKKTDPLFHPVETGVCKLNDNAQTDKAIEERLGGPLDVTVSQRYAK 188  
 Db 124 AGLAPDCAITVHTQADTSSPDPLFNPLKTVGCQLDNANVTDAILSRAGGSIAADFTGHRQT 183  
 QY 189 PPAQMGDVINFAASPYCKSLQOQKTCDFAHFAANEVNVNKEGTKVTLSGPLALSSTLGE 248  
 Db 184 AFRELERVLNFPQSNLCRKREKQDESCSLTQALPSELKVSAD--NVSLTGAVSLASMLTE 241  
 QY 249 IFLQNAQAMPEVAVQRLKGAENWVLSLSLHNAQNLMAKTPYIARHKGTPILLOQIDTAL 308  
 Db 242 IFLQQAQMGPEPGWGRITDHSQWNTLLSLHNAQFYLLQRTPEVARSATPLLDLKTAL 301  
 QY 309 T---LQDAQOGKLPISAQNRVFLFGGHDNTNANTAGMLGADWQLPQPDNTPPGGGLVF 365  
 Db 302 TPHPKQKAYGVTLPTS----VLFAGADTNTLANLGGALELNTWLPQGPDPNTPPGGELVF 357  
 QY 366 ELWQNPNDHQRVAVKMFYQTMQDLRNAEKLDKKNPAGIISVAVAGCE 414  
 Db 358 ERWRRLSDNSQWIVQSLVFQTLQQRDKTPLSL--NTPPGEVKLTLAGCE 405

RESULT 8  
 Q8XK29 PRELIMINARY; PRT; 432 AA.  
 ID Q8XK29  
 AC Q8XK29  
 DT 01-JUN-2002 (TREMELrel. 21, Created)  
 DT 01-JUN-2002 (TREMELrel. 21, Last sequence update)  
 DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)  
 DE Periplasmic phosphoanhydride phosphohydrolase.  
 GN APPA.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90368616; PubMed=2168385;  
 RA Dassa J., March C., Boquet P.-L.;  
 RT "The complete nucleotide sequence of the Escherichia coli gene appA  
 RT reveals significant homology between pH 2.5 acid phosphatase and  
 RT Glucose-1-phosphatase.";  
 RL J. Bacteriol. 172:5497-5500 (1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93054596; PubMed=1429631;  
 RA Ostanin K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.M.,  
 RA Van Etten R.L.;  
 RT "Overexpression, site-directed mutagenesis, and mechanism of  
 RT Escherichia coli acid phosphatase.";  
 RL J. Biol. Chem. 267:22830-22836 (1992).  
 DR EMBL; L03375; AAA00007.1;  
 DR GO; GO:0003993; F:acid phosphatase activity; IEA.  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR InterPro; IPR000560; HisAc\_phsphtse.  
 DR Pfam; PF00328; acid\_phosphat; 1.  
 DR PROSITE; PS00616; HIS\_ACID\_PHOSPHAT\_1; 1.  
 DR Hydrolyase.  
 SQ SEQUENCE 432 AA; 46990 MW; 951P393EA9A1A47C CRC64;

Query Match 42.2%; Score 929.5; DB 2; Length 432;  
 Best Local Similarity 47.7%; Pred. No. 4.2e-67;  
 Matches 195; Conservative 61; Mismatches 138; Indels 15; Gaps 7;  
 QY 13 LVLMSGLAAITAPVA---AEPSTYLERVILSRHGVSRPTKQTLQMDVTPDKWPWP 69  
 Db 5 LIPFLSLPLTPQSAFAQSEPE-LKLESVIVSRNGVRAPTKATQMLQMDVTPDAWPTWP 63

QY 70 VKAGVLTTPRGAEVLTLMGGFYGDYFRSLGLLA-AGCPAEGGVYAQADIDQRTLTGQAF 128  
 Db 64 VKLGLWLTTPRGELIAYLGHYQRLVADGLLAKKCPQSGQVAILADVDERTRKTGEAFA 123  
 QY 129 DGVAFCGLTVHNOADLKKTDPLFHPVETGVCKLNDNAQTDKAIEERLGGPLDVTVSQRYAK 188  
 Db 124 AGLAPDCAITVHTQADTSSPDPLFNPLKTVGCQLDNANVTDAILSRAGGSIAADFTGHRQT 183  
 QY 189 PPAQMGDVINFAASPYCKSLQOQKTCDFAHFAANEVNVNKEGTKVTLSGPLALSSTLGE 248  
 Db 184 AFRELERVLNFPQSNLCRKREKQDESCSLTQALPSELKVSAD--NVSLTGAVSLASMLTE 241  
 QY 249 IFLQNAQAMPEVAVQRLKGAENWVLSLSLHNAQNLMAKTPYIARHKGTPILLOQIDTAL 308  
 Db 242 IFLQQAQMGPEPGWGRITDHSQWNTLLSLHNAQFYLLQRTPEVARSATPLLDLKTAL 301  
 QY 309 T---LQDAQOGKLPISAQNRVFLFGGHDNTNANTAGMLGADWQLPQPDNTPPGGGLVF 365  
 Db 302 TPHPKQKAYGVTLPTS----VLFAGADTNTLANLGGALELNTWLPQGPDPNTPPGGELVF 357  
 QY 366 ELWQNPNDHQRVAVKMFYQTMQDLRNAEKLDKKNPAGIISVAVAGCE 414  
 Db 358 ERWRRLSDNSQWIVQSLVFQTLQQRDKTPLSL--NTPPGEVKLTLAGCE 405

RESULT 9  
 Q8XC29 PRELIMINARY; PRT; 434 AA.  
 ID Q8XC29  
 AC Q8XC29  
 DT 01-MAR-2002 (TREMELrel. 20, Created)  
 DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)  
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
 DE Phosphoanhydride phosphorylase, pH 2.5 acid phosphatase,  
 DE periplasmic.  
 GN APPA OR Z1397 OR ECS1136.  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoukis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";  
 RL Nature 409:529-533 (2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / RIMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:11-22 (2001).  
 DR EMBL; AB005292; AAG55528.1; ALT\_INIT.  
 DR EMBL; AP002554; BAB34559.1; -;  
 DR PIR; D85633; D85633.  
 DR PIR; H90770; H90770.  
 DR GO; GO:0003993; F:acid phosphatase activity; IEA.  
 DR InterPro; IPR000560; HisAc\_phsphtse.  
 DR Pfam; PF00328; acid\_phosphat; 1.  
 DR PROSITE; PS00616; HIS\_ACID\_PHOSPHAT\_1; 1.  
 DR PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.  
 KW Complete proteome.

SQ SEQUENCE 434 AA; 47337 MW; F197DFD1869F9C4 CRC64;

Query Match  
Best Local Similarity 42.1%; Score 926.5; DB 16; Length 434;  
Matches 194; Conservative 62; Mismatches 139; Indels 15; Gaps 7;

QY 13 LVLMUSGLAAITAPVA---AAPS-GYTLERVILSRHGVSPTKQTQMLNDVTPDKWPQW 68  
Db LIPFLSLIPLTPKSAFAQSEPE-LKESVIVSRHGVRAPTKATQMLQMDVTPDANPTWP 73

QY 5 LIPFLSLIPLTPKSAFAQSEPE-LKESVIVSRHGVRAPTKATQMLQMDVTPDANPTWP 64  
Db LIPFLSLIPLTPKSAFAQSEPE-LKESVIVSRHGVRAPTKATQMLQMDVTPDANPTWP 63

QY 69 PVKAGYLTPRGAELVTLMGGYGYDFRSLGLLA-AGCPAEGGVYVAQADIDORTLTGQAPL 127  
Db PVKAGYLTPRGAELVTLMGGYGYDFRSLGLLA-AGCPAEGGVYVAQADIDORTLTGQAPL 128

QY 55 PVKGLWLTFRGGLIAYLGHYQORLVADGLLTKGCPQGVQVAITADVDERTRKTGEAF 124  
Db PVKGLWLTFRGGLIAYLGHYQORLVADGLLTKGCPQGVQVAITADVDERTRKTGEAF 123

QY 128 LDGVPAGGGLTVHQADLTKTDPFLFHPVETGVCKLNAQTDKAEERLGGPLDVTVSQRYA 187  
Db LDGVPAGGGLTVHQADLTKTDPFLFHPVETGVCKLNAQTDKAEERLGGPLDVTVSQRYA 188

QY 125 AAGLAPDCAITVHTQADTSSDFLNFPLKTVGCVQCLDANVTDAILSRAGGSIAFDFTGHRQ 184  
Db AAGLAPDCAITVHTQADTSSDFLNFPLKTVGCVQCLDANVTDAILSRAGGSIAFDFTGHRQ 183

QY 188 KPFAGMGDLVNFAPSAFPCYSKLSQOQKTCDFPAHFAANEVNVNKGKTVTLSPGLALSITLG 247  
Db KPFAGMGDLVNFAPSAFPCYSKLSQOQKTCDFPAHFAANEVNVNKGKTVTLSPGLALSITLG 248

QY 185 TAFRELERVNFPOSNLCNREKQDECSLTQALPSELKVSAD--NVSLTGAVSLASMLT 242  
Db TAFRELERVNFPOSNLCNREKQDECSLTQALPSELKVSAD--NVSLTGAVSLASMLT 251

QY 248 EIFFLQNAQAMPVAVQORLKAENWVSLLSHNAQFNLMAKTYPIARHKGTPLLQOQIDTA 307  
Db EIFFLQNAQAMPVAVQORLKAENWVSLLSHNAQFNLMAKTYPIARHKGTPLLQOQIDTA 308

QY 243 EIFFLQNAQAMPVAVQORLKAENWVSLLSHNAQFNLMAKTYPIARHKGTPLLQOQIDTA 302  
Db EIFFLQNAQAMPVAVQORLKAENWVSLLSHNAQFNLMAKTYPIARHKGTPLLQOQIDTA 311

QY 308 LT---LQDAGQOKLPISAQNRVFLPGHDTNIAIAGMLGADWOLPEQPDNTPPGGLV 364  
Db LT---LQDAGQOKLPISAQNRVFLPGHDTNIAIAGMLGADWOLPEQPDNTPPGGLV 365

QY 303 LTPHPQQAQVGLTPS---VLFIAHDTNLANLGALNLTLPFGQPDNTPPGGLV 358  
Db LTPHPQQAQVGLTPS---VLFIAHDTNLANLGALNLTLPFGQPDNTPPGGLV 367

QY 355 FELQNPONHORYVAVKMFYQTMQDLNNAEKLDLKNPAGIISVAVAGCE 414  
Db FELQNPONHORYVAVKMFYQTMQDLNNAEKLDLKNPAGIISVAVAGCE 415

QY 359 FERWRSLDSNQWISQVSLVFTLQOQMDKTPLSL-NTFPGEVKLTLAGCE 407  
Db FERWRSLDSNQWISQVSLVFTLQOQMDKTPLSL-NTFPGEVKLTLAGCE 415

RESULT 10

Q8CW75 PRELIMINARY; PRT; 446 AA.

AC Q8CW75  
DT 01-WAR-2003 (TremBLrel. 23, Created)  
DT 01-WAR-2003 (TremBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)  
DE Periplasmic appa protein precursor.  
GN APPA OR C1121.  
OS Escherichia coli O6.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=217992;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;  
RX MEDLINE=22388234; PubMed=12471157;  
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,  
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Ferna N.T.,  
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
RT "Extensive mosaic structure revealed by the complete genome sequence  
of uropathogenic Escherichia coli.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).  
DR EMBL; AE016759; AAN79589.1; -  
DR GO; GO:0003993; F:acid phosphatase activity; IEA.  
DR InterPro; IPR000560; HisAc\_Phpase.  
DR Pfam; PF03228; acid\_phosphat; i.  
DR PROSITE; PS00616; HIS\_ACID\_PHOSPHAT\_1; 1.  
DR PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.  
KW Complete proteome.  
SQ SEQUENCE 446 AA; 48605 MW; F1308CD691DB9F49 CRC64;

Query Match  
Best Local Similarity 42.1%; Score 926.5; DB 16; Length 446;  
Matches 193; Conservative 63; Mismatches 138; Indels 15; Gaps 7;

QY 13 LVLMUSGLAAITAPVA---AAPS-GYTLERVILSRHGVSPTKQTQMLNDVTPDKWPQW 69  
Db LIPFLSLIPLTPKSAFAQSEPE-LKESVIVSRHGVRAPTKATQMLQMDVTPDANPTWP 73

Db LIPFLSLIPLTPKSAFAQSEPE-LKESVIVSRHGVRAPTKATQMLQMDVTPDANPTWP 73

QY 70 VKAGYLTPRGAELVTLMGGYGYDFRSLGLLA-AGCPAEGGVYVAQADIDORTLTGQAPL 128  
Db VKAGYLTPRGAELVTLMGGYGYDFRSLGLLA-AGCPAEGGVYVAQADIDORTLTGQAPL 128

QY 74 VKGLWLTFRGGLIAYLGHYQORLVADGLLTKGCPQGVQVAITADVDERTRKTGEAF 133  
Db VKGLWLTFRGGLIAYLGHYQORLVADGLLTKGCPQGVQVAITADVDERTRKTGEAF 123

QY 129 DGVPAGGGLTVHQADLTKTDPFLFHPVETGVCKLNAQTDKAEERLGGPLDVTVSQRYA 188  
Db DGVPAGGGLTVHQADLTKTDPFLFHPVETGVCKLNAQTDKAEERLGGPLDVTVSQRYA 188

QY 134 AGLAPDCAITVHTQADTSSDFLNFPLKTVGCVQCLDANVTDAILSRAGGSIAFDFTGHRQ 193  
Db AGLAPDCAITVHTQADTSSDFLNFPLKTVGCVQCLDANVTDAILSRAGGSIAFDFTGHRQ 193

QY 189 PFAQMGDLVNFAPSAFPCYSKLSQOQKTCDFPAHFAANEVNVNKGKTVTLSPGLALSITLG 248  
Db PFAQMGDLVNFAPSAFPCYSKLSQOQKTCDFPAHFAANEVNVNKGKTVTLSPGLALSITLG 251

QY 194 AFRLELERVNFPOSNLCNREKQDECSLTQALPSELKVSAD--NVSLTGAVSLASMLT 251  
Db AFRLELERVNFPOSNLCNREKQDECSLTQALPSELKVSAD--NVSLTGAVSLASMLT 251

QY 249 EIFFLQNAQAMPVAVQORLKAENWVSLLSHNAQFNLMAKTYPIARHKGTPLLQOQIDTA 308  
Db EIFFLQNAQAMPVAVQORLKAENWVSLLSHNAQFNLMAKTYPIARHKGTPLLQOQIDTA 311

QY 252 EIFFLQNAQAMPVAVQORLKAENWVSLLSHNAQFNLMAKTYPIARHKGTPLLQOQIDTA 311  
Db EIFFLQNAQAMPVAVQORLKAENWVSLLSHNAQFNLMAKTYPIARHKGTPLLQOQIDTA 311

QY 309 T---LQDAGQOKLPISAQNRVFLPGHDTNIAIAGMLGADWOLPEQPDNTPPGGLV 365  
Db T---LQDAGQOKLPISAQNRVFLPGHDTNIAIAGMLGADWOLPEQPDNTPPGGLV 367

QY 312 TFPSPQQAQVGLTPS---VLFIAHDTNLANLGALNLTLPFGQPDNTPPGGLV 367  
Db TFPSPQQAQVGLTPS---VLFIAHDTNLANLGALNLTLPFGQPDNTPPGGLV 367

QY 366 ELQNPONHORYVAVKMFYQTMQDLNNAEKLDLKNPAGIISVAVAGCE 414  
Db ELQNPONHORYVAVKMFYQTMQDLNNAEKLDLKNPAGIISVAVAGCE 415

QY 368 ERWRSLDSNQWISQVSLVFTLQOQMDKTPLSL-NTFPGEVKLTLAGCE 415  
Db ERWRSLDSNQWISQVSLVFTLQOQMDKTPLSL-NTFPGEVKLTLAGCE 415

RESULT 11

Q7UD08 PRELIMINARY; PRT; 432 AA.

AC Q7UD08  
DT 01-OCT-2003 (TremBLrel. 25, Created)  
DT 01-OCT-2003 (TremBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)  
DE Phosphoanhydride phosphorylase; pH 2.5 acid phosphatase.  
GN APPA OR S1048.  
OS Shigella flexneri.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Shigella.  
OX NCBI\_TaxID=623;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=2457T / ATCC 700930 / Serotype 2a;  
RX MEDLINE=22590274; PubMed=12704152;  
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,  
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,  
RA Mau B., Ferna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,  
RA Schwartz D.C., Blattner F.R.;  
RT "Complete genome sequence and comparative genomics of Shigella  
flexneri serotype 2a strain 2457T.";  
RL Infect. Immun. 71:2775-2786 (2003).  
DR EMBL; AE016981; AAP16495.1; -  
SQ SEQUENCE 432 AA; 47136 MW; 200442239BC66DFC CRC64;

Query Match  
Best Local Similarity 41.8%; Score 919.5; DB 16; Length 432;  
Matches 193; Conservative 63; Mismatches 138; Indels 15; Gaps 7;

QY 13 LVLMUSGLAAITAPVA---AAPS-GYTLERVILSRHGVSPTKQTQMLNDVTPDKWPQW 69  
Db LIPFLSLIPLTPKSAFAQSEPE-LKESVIVSRHGVRAPTKATQMLQMDVTPDANPTWP 73

QY 70 VKAGYLTPRGAELVTLMGGYGYDFRSLGLLA-AGCPAEGGVYVAQADIDORTLTGQAPL 128  
Db VKAGYLTPRGAELVTLMGGYGYDFRSLGLLA-AGCPAEGGVYVAQADIDORTLTGQAPL 128

QY 74 VKGLWLTFRGGLIAYLGHYQORLVADGLLTKGCPQGVQVAITADVDERTRKTGEAF 123  
Db VKGLWLTFRGGLIAYLGHYQORLVADGLLTKGCPQGVQVAITADVDERTRKTGEAF 123

QY 129 DGVPAGGGLTVHQADLTKTDPFLFHPVETGVCKLNAQTDKAEERLGGPLDVTVSQRYA 188  
Db DGVPAGGGLTVHQADLTKTDPFLFHPVETGVCKLNAQTDKAEERLGGPLDVTVSQRYA 188

QY 124 AGLAPDCAITVHTQADTSSDFLNFPLKTVGCVQCLDANVTDAILSRAGGSIAFDFTGHRQ 183  
Db AGLAPDCAITVHTQADTSSDFLNFPLKTVGCVQCLDANVTDAILSRAGGSIAFDFTGHRQ 183

QY 189 PFAQMGDLVNFAPSAFPCYSKLSQOQKTCDFPAHFAANEVNVNKGKTVTLSPGLALSITLG 248  
Db PFAQMGDLVNFAPSAFPCYSKLSQOQKTCDFPAHFAANEVNVNKGKTVTLSPGLALSITLG 248

184 VFERLRLVNFPOSNLCNREKQSCSLTQALPSELKVSAD--NVSLTGAVSLASMLTE 241  
249 IFLLQQAQMPVAVQWKLKGAENWVSLSLHNAQFNLMKTPYIAHKGTPPLQQIDTAL 308  
242 IFLLQQAQMPVAVQWKLKGAENWVSLSLHNAQFNLMKTPYIAHKGTPPLQQIDTAL 301  
309 T---LQIDAQCKLPISAQNRVFLGGHDTNINAGMGADWOLPEQDNTPPGGGLVF 365  
302 TPFPQKQAYGVTLPTS---VLFIAGHDTNLANLGGALELNWTLPGQDNTPPGGGLVF 357  
366 ELWQPNPDHQRVAVKMFYQTMQDLRQNAEKLDLKNPAGIISVAVAGCE 414  
358 ERWRSLSDNSQWISQVSLVFTLQQRDKTPLSL--NTPPGEVKLTLAGCE 405

RESULT 12  
Q83RW2 PRELIMINARY; PRT; 442 AA.  
ID Q83RW2 AC Q83RW2  
DT 01-JUN-2003 (TREMBLrel. 24, Created)  
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Phosphoanhydride phosphorilase, pH 2.5 acid phosphatase,  
DE periplasmic  
GN APFA OR SF0982.  
OS Shigella flexneri.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Shigella.  
NCBI\_TaxID=623;  
RN NCBI\_TaxID=623;  
RP SEQUENCE FROM N.A.  
RC STRAIN=301 / Serotype 2a; PubMed=12384590;  
RX MEDLINE=2227406; PubMed=12384590;  
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,  
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,  
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,  
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,  
RA Yu J.,  
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity  
RT through comparison with genomes of Escherichia coli K12 and O157.",  
RL Nucleic Acids Res. 30:4432-4441 (2002).  
DR EMBL; A015127; AAN42610.1; -  
DR GO; GO:0003993; F:acid phosphatase activity; IEA.  
DR InterPro; IPR000560; HisAc\_phsphtse.  
DR Pfam; PF00328; acid phosphat; 1.  
DR PROSITE; PS00616; HIS ACID PHOSPHAT 1; 1.  
DR PROSITE; PS00778; HIS ACID PHOSPHAT 2; 1.  
KW Complete proteome.  
SQ SEQUENCE 442 AA; 48183 MW; 20BCE2C454254680 CRC64;

Query Match 41.8%; Score 919.5; DB 16; Length 442;  
Best Local Similarity 47.2%; Pred. No. 2.8e-66;  
Matches 193; Conservative 63; Mismatches 138; Indels 15; Gaps 7;

13 LVMLSGLAATAPVA---AEPGTYLERVLSRHGVRSPKQTLQNDVTPDKWPWP 69  
15 LIPFLSLIPLTPQSAFAQSEPE-LKLESMTVSRHGVRAPTKATQLMQDVTPDAMPWP 73

70 VKAGYLTTPRGAELVTLMGFGYDGFPSLGLLA-AGCPAEGGYAQAADIDQRTLTQAF 128  
74 VKLGMWLTTPRGGELVTLMGFGYDGFPSLGLLA-AGCPAEGGYAQAADIDQRTLTQAF 133

129 DGVAQCGGLTVHQAADLKKTDPLFHPVETGVCKLQNAQDPAIEERLGGPLDVTQVAK 188  
134 AGLAPDCAITVHTQADTSSPDPLFNPLKTVGVQQLDNANVTAILCRAGSIAADFTGHRQT 193

189 PFAQMGDLNFAASPKSLQOQKTCDFAHFAAENVNVNKGSKVTLGSLALSTLGE 248  
194 VFRELRLVNFPOSNLCNREKQSCSLTQALPSELKVSAD--NVSLTGAVSLASMLTE 251

249 IFLLQQAQMPVAVQWKLKGAENWVSLSLHNAQFNLMKTPYIAHKGTPPLQQIDTAL 308  
252 IFLLQQAQMPVAVQWKLKGAENWVSLSLHNAQFNLMKTPYIAHKGTPPLQQIDTAL 311

309 T---LQIDAQCKLPISAQNRVFLGGHDTNINAGMGADWOLPEQDNTPPGGGLVF 365  
312 TPFPQKQAYGVTLPTS---VLFIAGHDTNLANLGGALELNWTLPGQDNTPPGGGLVF 367  
366 ELWQPNPDHQRVAVKMFYQTMQDLRQNAEKLDLKNPAGIISVAVAGCE 414  
368 ERWRSLSDNSQWISQVSLVFTLQQRDKTPLSL--NTPPGEVKLTLAGCE 415

RESULT 13  
Q8XBZ6 PRELIMINARY; PRT; 413 AA.  
ID Q8XBZ6 AC Q8XBZ6  
DT 01-MAR-2002 (TREMBLrel. 20, Created)  
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
DE Periplasmic glucose-1-phosphatase.  
GN AGP OR Z1421 OR ECS1158.  
OS Escherichia coli O157:H7.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
NCBI\_TaxID=83334;  
RN NCBI\_TaxID=83334;  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
RX MEDLINE=21074935; PubMed=11206551;  
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
RA Postai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocis K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blattner F.R.,  
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.",  
RL Nature 409:529-533 (2001).  
RN NCBI\_TaxID=83334;  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / RIMD 050952;  
RX MEDLINE=21156231; PubMed=11258796;  
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
RA Kunara S., Shiba T., Hattori M., Shingawa H.,  
RA RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
RT O157:H7 and genomic comparison with a laboratory strain K-12.",  
RL DNA Res. 8:11-22 (2001).  
DR EMBL; A005294; AAG55550.1; -  
DR EMBL; AP002554; BAB34581.1; -  
DR PIR; B85636; B85636.  
DR PIR; P90773; P90773.  
DR GO; GO:0003993; F:acid phosphatase activity; IEA.  
DR InterPro; IPR000560; HisAc\_phsphtse.  
DR Pfam; PF00328; acid phosphat; 1.  
DR PROSITE; PS00616; HIS ACID PHOSPHAT 1; 1.  
DR PROSITE; PS00778; HIS ACID PHOSPHAT 2; 1.  
KW Complete proteome.  
SQ SEQUENCE 413 AA; 45640 MW; A4630644EF66FB1A CRC64;

Query Match 24.2%; Score 532; DB 16; Length 413;  
Best Local Similarity 31.3%; Pred. No. 9e-35;  
Matches 131; Conservative 83; Mismatches 164; Indels 40; Gaps 11;

10 LSGVLTTPRGAELVTLMGFGYDGFPSLGLLA-AGCPAEGGYAQAADIDQRTLTQAF 68  
10 VAGVLTTPRGAELVTLMGFGYDGFPSLGLLA-AGCPAEGGYAQAADIDQRTLTQAF 64

69 PVKAGYLTTPRGAELVTLMGFGYDGFPSLGLLA-AGCPAEGGYAQAADIDQRTLTQAF 127  
65 DVPGQGLTKGVLVTLMGFGYDGFPSLGLLA-AGCPAEGGYAQAADIDQRTLTQAF 124

128 LDGVAQCGGLTVHQAADLKKTDPLFHPVETGVCKLQNAQDPAIEERLGGPLDVTQVAK 187  
125 ITGAPPGCDIPVHQAQMPVAVQWKLKGAENWVSLSLHNAQFNLMKTPYIAHKGTPPLQQIDTAL 178

```
QY 188 KPFAQMGDVLFNAASPYCKSLQ-----QKTCDFAHFAANEVNVNKEGTVKTLTGQAF 243
Db 179 DSYQLLEKIVNYKDSPACKCKQCSLVDGKNT-----FSA-----KYQEPGVSGPLKVG 228
QY 244 STIGEIFLLQNAQAMP--EVAWQRLKGAENWVSLSLHNAQFNLMAKTPYIARHKGTPLL 301
Db 229 NSLVDFTLOYEGFMDQVANGKISDQWKVSLKNGYQDSLFTSPEVARNVAKPLV 288
QY 302 QOITDALTQLDQAQOKLPISAQNRVFLFGGHDNTNIAAGMLG-ADWQLPEQPDNTPPG 360
Db 289 SYIDKALVTDR-----TSAPKITVLVGHDSNIAASLTALDFKPYQLHDQNTPTIG 339
QY 361 GGLVFLWQPNPDHORYVAVKMFYQTMQDLRNEAKLDLKNPAGIISVAVAGC--ENNG 417
Db 340 GKIVFQRWDSKANRDLMKIEYVYQSAQLRNDALTQ-APAQRVTLSELSGCPIDAD 405

RESULT 14
Q7UD02 PRELIMINARY; PRT; 413 AA.
AC Q7UD02; 2003 (TREMBlurel. 25, Created)
DT 01-OCT-2003 (TREMBlurel. 25, last sequence update)
DE Periplasmic glucose-1-phosphatase.
GN AGP OR S1075.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Ferna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
RT flexneri serotype 2a strain 2457T."
RL Infect. Immun. 71:2775-2786(2003).
DR EMBL; AE016981; AAP16516.1; -.
SQ SEQUENCE 413 AA; 45648 MW; 6996CAE70CC65B39 CRC64;

Query Match 24.0%; Score 527; DB 16; Length 413;
Best Local Similarity 31.3%; Pred. No. 2.3e-34;
Matches 131; Conservative 83; Mismatches 163; Indels 42; Gaps 12;

QY 10 LSGVLMLSLGLAATAPVAAPSGYTLERVVILSRHGVRSP-TKQTOLMNDVTPDKWPQW 68
Db 10 VAGIVLASNAQAQTV-----PEGYQLQVLMMSRHLRAPLANNGSVLEQSTFNKWPW 64
QY 69 PVKAGYLTTPRGAEVLTMGGFYGYFRSLGLLAAG-CPAEGGVYQAQIDQRTLTGQAF 127
Db 65 DVPQGQLTTKGVLEVYMGHYMREWLAEQGIKSGECPPTTVYVANSIQRVATAQFF 124
QY 128 LDGVAPGGLTVHQAQDLKKTDPFLFHPVETGVCCKLDAQTDKALIEERLGGPLDVTVSORYA 187
Db 125 ITGAPFGCDIPVHQRKMGTMWPTFNPMITDSSAFSEQAVAAMEKELS-----KLQLT 178
QY 188 KPFAQMGDVLFNAASPYCKSLQ-----QKTCDFAHFAANEVNVNKEGTVKTLTGQAF 243
Db 179 DSYQLLEKIVNYKDSPACKCKQCSLVDGKNT-----FSA-----KYQEPGVSGPLKVG 228
QY 244 STIGEIFLLQNAQAMP--EVAWQRLKGAENWVSLSLHNAQFNLMAKTPYIARHKGTPLL 301
Db 229 NSLVDFTLOYEGFMDQVANGKISDQWKVSLKNGYQDSLFTSPEVARNVAKPLV 288
QY 302 QOITDALTQLDQAQOKLPISAQNRVFLFGGHDNTNIAAGMLG-ADWQLPEQPDNTPPG 360
Db 289 SYIDKALVTDR-----TSAPKITVLVGHDSNIAASLTALDFKPYQLHDQNTPTIG 339
```

```
QY 361 GGLVFLWQPNPDHORYVAVKMFYQTMQDLRNEAKLDLKNPAGIISVAVAGC--ENNG 417
Db 340 GKIVFQRWDSKANRDLMKIEYVYQSAQLRNDALTQ-APAQRVTLSELSGCPIDANG 397

RESULT 15
Q83RV6 PRELIMINARY; PRT; 421 AA.
AC Q83RV6; 2003 (TREMBlurel. 24, Created)
DT 01-JUN-2003 (TREMBlurel. 24, last sequence update)
DE Periplasmic glucose-1-phosphatase.
GN AGP OR SF1006.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=301 / Serotype 2a;
RX MEDLINE=2272406; PubMed=12384590;
RA Jin O., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157."
RL Nucleic Acids Res. 30:4432-4441(2002).
DR EMBL; AE015129; AAN42632.1; -.
DR GO; GO:0003993; P.acid phosphatase activity; IEA.
DR InterPro; IPR000560; HisAc_phosphatase.
DR Pfam; PF00328; acid_phosphat; 1.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
KW Complete proteome.
SQ SEQUENCE 421 AA; 46504 MW; 6D702665F3E8B6FC CRC64;

Query Match 24.0%; Score 527; DB 16; Length 421;
Best Local Similarity 31.3%; Pred. No. 2.4e-34;
Matches 131; Conservative 83; Mismatches 163; Indels 42; Gaps 12;

QY 10 LSGVLMLSLGLAATAPVAAPSGYTLERVVILSRHGVRSP-TKQTOLMNDVTPDKWPQW 68
Db 18 VAGIVLASNAQAQTV-----PEGYQLQVLMMSRHLRAPLANNGSVLEQSTFNKWPW 72
QY 69 PVKAGYLTTPRGAEVLTMGGFYGYFRSLGLLAAG-CPAEGGVYQAQIDQRTLTGQAF 127
Db 73 DVPQGQLTTKGVLEVYMGHYMREWLAEQGIKSGECPPTTVYVANSIQRVATAQFF 132
QY 128 LDGVAPGGLTVHQAQDLKKTDPFLFHPVETGVCCKLDAQTDKALIEERLGGPLDVTVSORYA 187
Db 133 ITGAPFGCDIPVHQRKMGTMWPTFNPMITDSSAFSEQAVAAMEKELS-----KLQLT 186
QY 188 KPFAQMGDVLFNAASPYCKSLQ-----QKTCDFAHFAANEVNVNKEGTVKTLTGQAF 243
Db 187 DSYQLLEKIVNYKDSPACKCKQCSLVDGKNT-----FSA-----KYQEPGVSGPLKVG 236
QY 244 STIGEIFLLQNAQAMP--EVAWQRLKGAENWVSLSLHNAQFNLMAKTPYIARHKGTPLL 301
Db 237 NSLVDFTLOYEGFMDQVANGKISDQWKVSLKNGYQDSLFTSPEVARNVAKPLV 296
QY 302 QOITDALTQLDQAQOKLPISAQNRVFLFGGHDNTNIAAGMLG-ADWQLPEQPDNTPPG 360
Db 297 SYIDKALVTDR-----TSAPKITVLVGHDSNIAASLTALDFKPYQLHDQNTPTIG 347
QY 361 GGLVFLWQPNPDHORYVAVKMFYQTMQDLRNEAKLDLKNPAGIISVAVAGC--ENNG 417
Db 348 GKIVFQRWDSKANRDLMKIEYVYQSAQLRNDALTQ-APAQRVTLSELSGCPIDANG 405

Search completed: April 30, 2004, 12:41:24
```

Job time : 41.0488 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 12:35:05 ; Search time 15 Seconds  
(without alignments)  
2629.235 Million cell updates/sec

Title: US-10-021-723B-13

Perfect score: 2153

Sequence: 1 QSEPELKLESVVIVSRHGVR.....CSLAGFTQIVNEARIPACSL 410

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78.\*

1: Piri.\*

2: Pir2.\*

3: Pir3.\*

4: Pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2077	96.5	432	2 B36733	acid phosphatase (
2	2042	94.8	434	2 H90770	phosphoanhydride p
3	2042	94.8	444	2 D85633	hypothetical prote
4	954	44.3	441	2 AC0201	acid phosphatase (
5	551	25.6	413	2 F90773	periplasmic glucos
6	551	25.6	413	2 B85636	periplasmic glucos
7	544	25.3	413	2 JV0087	glucose-1-phosphat
8	543	25.2	413	2 AG0632	glucose-1-phosphat
9	534.5	24.8	414	2 E87316	periplasmic phosph
10	518	24.1	417	2 S25627	glucose-1-phosphat
11	133	6.2	416	2 T16058	hypothetical prote
12	132	6.1	423	1 S06167	acid phosphatase (
13	114	5.3	423	2 A33395	acid phosphatase (
14	112.5	5.2	438	2 S84682	acid phosphatase (
15	109.5	5.1	344	2 E89130	paired box transcr
16	109.5	5.1	457	2 A56925	paired box transcr
17	109.5	5.1	479	1 JN0890	acid phosphatase (
18	107.5	5.0	459	2 S52250	paired box transcr
19	107.5	5.0	479	1 JN0715	3-phytase (EC 3.1.
20	107	5.0	421	2 S14742	acid phosphatase (
21	107	5.0	5126	2 S40450	ryanodine receptor
22	105	4.9	450	2 A84429	paired box transcr
23	104.5	4.9	397	2 C91716	hypothetical prote
24	103.5	4.8	537	2 S54770	secreted acid phos
25	103.5	4.8	888	2 T46726	secreted acid phos
26	103	4.8	769	2 B87681	tyrosine kinase Di
27	103	4.8	1048	1 BVCECSC	exonuclease (EC 3
28	101.5	4.7	356	2 P75594	hypothetical prote
29	101.5	4.7	386	1 JN0610	acid phosphatase (

#### RESULT 1

B36733

acid phosphatase (EC 3.1.3.2) precursor - Escherichia coli (strain K-12)  
N;Alternate names: phosphoanhydride phosphohydrolase, periplasmic; phytase 6  
C;Species: Escherichia coli

C;Date: 19-Apr-1991 #sequence revision 19-Apr-1991 #text change 01-Mar-2002

C;Accession: B36733; S18018; B64839; A36534; S17960; S33278

R;Dassa, J.; Marck, C.; Boquet, P.L.

J. Bacteriol. 172, 5497-5500, 1990

A;Title: The complete nucleotide sequence of the Escherichia coli gene appA reveals si

A;Reference number: A36733; MUID:90368616; PMID:2168385

A;Accession: B36733

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-432 <DAS>

A;Cross-references: GB:M58708; NID:9145283; PIDN:AAA72086.1; PID:9145285

R;Greiner, R.; Jany, K.D.

Biol. Chem. Hoppe-Seyler 372, 664-665, 1991

A;Title: Characterization of a phytase from Escherichia coli.

A;Reference number: S18018

A;Accession: S18018

A;Molecule type: protein

A;Residues: 23-33 <GRE>

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A;Title: The complete genome sequence of Escherichia coli K-12.

A;Reference number: A64720; MUID:97426617; PMID:9278503

A;Accession: B64839

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-432 <BLAT>

A;Cross-references: GB:AE000200; GB:U00096; NID:92367111; PIDN:AACT4065.1; PID:9178721

A;Experimental source: strain K-12, substrain MG1655

R;Touati, E.; Danchin, A.

Biochimie 69, 215-221, 1987

A;Title: The structure of the promoter and amino terminal region of the pH 2.5 acid ph

A;Reference number: A28534; MUID:87271766; PMID:3038201

A;Accession: A28534

A;Molecule type: DNA

A;Residues: 1-50, 'NAGCHPRMANLAG', 65, 'T', 67-74, 'DV', 77-111, 'S', <TOU>

A;Cross-references: GB:X05471; NID:940325; PIDN:CAA29031.1; PID:940927

R;Dassa, J.; Faisli, H.; Marck, C.; Dion, M.; Kieffer-Bontemps, M.; Boquet, P.L.

Mol. Gen. Genet. 229, 341-352, 1991

A;Title: A new oxygen-regulated operon in Escherichia coli comprises the genes for a p

A;Reference number: S17958; MUID:92049231; PMID:1658595

A;Accession: S17960

A;Status: not compared with conceptual translation

A;Molecule type: DNA

A;Residues: 1-17 <DA2>

A;Cross-references: GB:S63811; NID:9238656; PIDN:AAB20286.1; PID:9238659

R;Greiner, R.; Konietzny, U.; Jany, K.D.

Arch. Biochem. Biophys. 303, 107-113, 1993  
A>Title: Purification and characterization of two phytases from *Escherichia coli*.  
A/Reference number: S33278; MUID:93256556; PMID:8387749  
A/Accession: S33278  
A/Molecule type: protein  
A/Residues: 23-31, 'A', 33-34 <GR2>  
C/Comment: In addition to CAMP-mediated control, this enzyme is induced when bacterial  
naerobic conditions.  
C/Genetics:  
A/Gene: appA  
A/Map position: 25 min  
C/Keywords: monomer; periplasmic space; phosphohistidine; phosphoprotein; phosphoric mon  
F/1-22/Domain: signal sequence #status predicted <SIG>  
F/23-112/Product: acid phosphatase #status predicted <MAT>  
F/38/Active site: Arg #status predicted  
F/39/Active site: His (phosphohistidine intermediate) #status predicted

Query Match 96.5%; Score 2077; DB 2; Length 432;  
Best Local Similarity 99.0%; Pred. No. 4.7e-160;  
Matches 402; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 QSEPELKESVIVSRHGVRAPTKATQMODVTPDAMPTWPKLGELTPRGGELIAYLGH 60  
DB 23 QSEPELKESVIVSRHGVRAPTKATQMODVTPDAMPTWPKLGELTPRGGELIAYLGH 82  
QY 61 YWRQRLVADGLLPKCGCPQSGQVVAIIADVDERTRKTEGEAFAAGLAPDCAITVHTQADTSS 120  
DB 83 YQORQLVADGLLTKKGGCPQSGQVVAIIADVDERTRKTEGEAFAAGLAPDCAITVHTQADTSS 142  
QY 121 PDPLFNPDKTGVCGOLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLPQSNLCLK 180  
DB 143 PDPLFNPDKTGVCGOLDNANVTDAILSRAGSIADFTGHRQTAFRELERVLPQSNLCLK 202  
QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMPGPGWGRITDS 240  
DB 203 REKQDESCSLTQALPSELKVSADNVSLTGAVSLASMLTEIFLLQQAQMPGPGWGRITDS 262  
QY 241 HOWNTLISLHNAQFDLQRTPEVARSRATPLDLIKTALTPHPPOKQAYGVTLPTSVLFI 300  
DB 263 HOWNTLISLHNAQFYLLQRTPEVARSRATPLDLIKTALTPHPPOKQAYGVTLPTSVLFI 322  
QY 301 AGHDTNLANLGGALLENLWTLPGQPDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 360  
DB 323 AGHDTNLANLGGALLENLWTLPGQPDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 382  
QY 361 RDKTPLSLNTPPGGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 410  
DB 383 RDKTPLSLNTPPGGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 432

RESULT 2  
H90770  
phosphoanhydride phosphorylase [imported] - *Escherichia coli* (strain O157:H7, substrain  
C/Species: *Escherichia coli*  
C/Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A>Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gen  
A/Reference number: A99629; MUID:21156231; PMID:11258796  
A/Accession: H90770  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-434 <HAY>  
A/Cross-references: GB:BA000007; PIDN:BA034559.1; PID:G13360596; GSPDB:GN00154  
A/Experimental source: strain O157:H7, substrain RMD 0509952  
C/Genetics:  
A/Gene: ECs1136

Query Match 94.8%; Score 2042; DB 2; Length 434;  
Best Local Similarity 96.3%; Pred. No. 3.2e-157;  
Matches 395; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

QY 1 QSEPELKESVIVSRHGVRAPTKATQMODVTPDAMPTWPKLGELTPRGGELIAYLGH 60  
DB 25 EPEPELKESVIVSRHGVRAPTKATQMODVTPDAMPTWPKLGELTPRGGELIAYLGH 84  
QY 61 YWRQRLVADGLLPKCGCPQSGQVVAIIADVDERTRKTEGEAFAAGLAPDCAITVHTQADTSS 120  
DB 85 YQORQLVADGLLTKKGGCPQSGQVVAIIADVDERTRKTEGEAFAAGLAPDCAITVHTQADTSS 144  
QY 121 PDPLFNPDKTGVCGOLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLPQSNLCLK 180  
DB 145 PDPLFNPDKTGVCGOLDNANVTDAILSRAGSIADFTGHRQTAFRELERVLPQSNLCLK 204  
QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMPGPGWGRITDS 240  
DB 205 REKQDESCSLTQALPSELKVSADNVSLTGAVSLASMLTEIFLLQQAQMPGPGWGRITDS 264  
QY 241 HOWNTLISLHNAQFDLQRTPEVARSRATPLDLIKTALTPHPPOKQAYGVTLPTSVLFI 300  
DB 265 HOWNTLISLHNAQFYLLQRTPEVARSRATPLDLIKTALTPHPPOKQAYGVTLPTSVLFI 324  
QY 301 AGHDTNLANLGGALLENLWTLPGQPDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 360  
DB 325 AGHDTNLANLGGALLENLWTLPGQPDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 384  
QY 361 RDKTPLSLNTPPGGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 410  
DB 385 RDKTPLSLNTPPGGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 434

RESULT 3  
D85633  
hypothetical protein appA [imported] - *Escherichia coli* (strain O157:H7, substrain EDI  
C/Species: *Escherichia coli*  
C/Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayh  
iller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodac  
Nature 409, 529-533, 2001  
A>Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.  
A/Reference number: A85480; MUID:21074935; PMID:11206551  
A/Accession: D85633  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-444 <STO>  
A/Cross-references: GB:AE005174; NID:G12514245; PIDN:AA055528.1; GSPDB:GN00145; UMGPI:2  
A/Experimental source: strain O157:H7, substrain EDL933  
C/Genetics:  
A/Gene: appA

Query Match 94.8%; Score 2042; DB 2; Length 444;  
Best Local Similarity 96.3%; Pred. No. 3.3e-157;  
Matches 395; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

QY 1 QSEPELKESVIVSRHGVRAPTKATQMODVTPDAMPTWPKLGELTPRGGELIAYLGH 60  
DB 35 EPEPELKESVIVSRHGVRAPTKATQMODVTPDAMPTWPKLGELTPRGGELIAYLGH 94  
QY 61 YWRQRLVADGLLPKCGCPQSGQVVAIIADVDERTRKTEGEAFAAGLAPDCAITVHTQADTSS 120  
DB 95 YQORQLVADGLLTKKGGCPQSGQVVAIIADVDERTRKTEGEAFAAGLAPDCAITVHTQADTSS 154  
QY 121 PDPLFNPDKTGVCGOLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLPQSNLCLK 180  
DB 155 PDPLFNPDKTGVCGOLDNANVTDAILSRAGSIADFTGHRQTAFRELERVLPQSNLCLK 214  
QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMPGPGWGRITDS 240  
DB 215 REKQDESCSLTQALPSELKVSADNVSLTGAVSLASMLTEIFLLQQAQMPGPGWGRITDS 274  
QY 241 HOWNTLISLHNAQFDLQRTPEVARSRATPLDLIKTALTPHPPOKQAYGVTLPTSVLFI 300  
DB 275 HOWNTLISLHNAQFYLLQRTPEVARSRATPLDLIKTALTPHPPOKQAYGVTLPTSVLFI 334

301 AGHDTNLANLGALELNTLPGQDNTPTPGGELVFERWRRLSDNSQWISQVLTQLOOM 360  
 335 AGHDTNLANLGALELNTLPGQDNTPTPGGELVFERWRRLSDNSQWISQVLTQLOOM 394  
 361 RDKTFLSLNTTGGVVKLTLAGCEERNAQMGCSLAGTQIVNEARIPACSL 410  
 395 RDKTFLSLNTTGGVVKLTLAGCEERNAQMGCSLAGTQIVNEARIPACSL 444

RESULT 4  
 AC0201  
 acid phosphatase (EC 3.1.3.2) [imported] - Yersinia pestis (strain CO92)  
 C:Species: Yersinia pestis  
 C>Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 27-Nov-2001  
 C:Accession: AC0201  
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Hoiden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; deo-Tarraga, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, N.; M.; Rutherford, K.; 523-527, 2001  
 Nature 413, 523-527, 2001  
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
 A:Reference number: AB0001; MUID:21470413; PMID:11586360  
 A:Accession: AC0201  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-441 <KUR>  
 A:Cross-references: GB:AL590842; PIDN:CAC90470.1; PID:gl5979685; GSPDB:GN00175  
 A:Gene: YPO1648  
 C:Genetics:  
 C:Keywords: phosphoric monoester hydrolase

Query Match 44.3%; Score 954; DB 2; Length 441;  
 Best Local Similarity 46.8%; Pred. No. 2.9e-69;  
 Matches 195; Conservative 65; Mismatches 145; Indels 12; Gaps 6;  
 2 SEPE-LKLESVIVSRHGVRAPTKATQMLMDVTPDAWPTWPKLGELTPRGGLIAYLGH 60  
 29 APSGYTLERVVLSRHGVRSPTKQTLMDVTPDKWPQWPKAGYLLTPRGGLIAYLGH 88  
 61 YRQRLVADGLPKKCPQSGQVIAIADVDERTKGTGEAFAGLAPDCAITVHTQADTSS 120  
 89 FGDYFRSLGLL-AAGCPAEGGVYACADIDQRLTGTGAFDGVAPGGLTVHQAADLKK 147  
 121 PDPLFNPLKTGVCOLDNANVTDAILERAGGSIAFTGHYQTAFARELVNPPQSNCLK 180  
 148 TDFLPHFVAVGCKLDAADTKAIEBQLGGLDTSQRYAKPFAQMGVLYNFAASPYCKS 207  
 181 REKQDSCSLTQALPSELKVSAD-CVSLTGAVSLASMLTEIFLLQQAQMGPEPCWGRIT 238  
 208 LQQQGGTKDFAHFAANEVNVNKGKTVLUSGLAUSSTLGEIFLQNAQMEVAVQRLK 267  
 239 DSHQWNTLLSLHNAQFDLQRTPEVARSRATFLDLIKTALTTPHPKQAYGVTLPTS-- 296  
 268 GAENWVSLLSLHNAQFNLAQTPYIARHKGTLPLQIDTALT--LQLDAQOKLPISAQ 324  
 297 --VLFTAGHDTNLANLGALELNTLPGQDNTPTPGGELVFERWRRLSDNSQWISQV 354  
 325 NRVFLTGGHDTNLANLGALELNTLPGQDNTPTPGGELVFERWRRLSDNSQWISQV 384  
 355 QTLQOMRDKTFPLSL-NTPPGEVKLTLAGCEERNAQMGCSLAGTQIVNEARIPACSL 410  
 385 QTMQDLNRAEKLDIKNNPAGIISVAVAGCENNGDKLCELDITFQKKVAVIEPACHI 441

RESULT 5  
 F90773  
 periplasmic glucose-1-phosphatase [imported] - Escherichia coli (strain O157:H7, substra  
 C:Species: Escherichia coli  
 C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
 C:Accession: F90773  
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
 gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and ge  
 A:Reference number: A99629; MUID:21156231; PMID:11258796  
 A:Accession: F90773  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-413 <HAY>  
 A:Cross-references: GB:BA000007; PIDN:BA034581.1; PID:gl3360618; GSPDB:GN00154  
 A:Experimental source: strain O157:H7, substrain RMD 0509952  
 C:Genetics:  
 A:Gene: ECs1158

Query Match 25.6%; Score 551; DB 2; Length 413;  
 Best Local Similarity 33.3%; Pred. No. 1e-36;  
 Matches 138; Conservative 71; Mismatches 169; Indels 36; Gaps 12;  
 1 QSEPE-LKLESVIVSRHGVRAPTKATQMLMDVTPDAWPTWPKLGELTPRGGLIAYL 58  
 23 QTVPEGYQLQCVLMMSRHLNRLAPLANNVSVLSQSTPNKWPEDVPGGQLTTKGGVLEVM 82  
 59 GHYWRQLVADGLPKKCPQSGQVIAIADVDERTKGTGEAFAGLAPDCAITVHTQADT 118  
 83 GHYWRQLVADGLPKKCPQSGQVIAIADVDERTKGTGEAFAGLAPDCAITVHTQADT 142  
 119 SSPDPLFNPLKTGVCOLDNANVTDAILERAGGSIAFTGHYQTAFARELVNPPQSN 175  
 143 GTWDPFTFNPTVD---DAAAFSEQAAMAEKLSKL-QLTDSYQL-----LEKIVNKDS 193  
 176 NLCKEKESKDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMGPE--EPG 233  
 194 PAC-----XKQKQCSLVGDKNTPSAKYQBPVSGPLKVGNSLVDAFTLQVYEGFPMDOVA 249  
 234 WGRITDSHONTLLSLHNAQFDLQRTPEVARSRATFLDLIKTALTTPHPKQAYGVTL 293  
 250 WGBIKSDQKQVLSKLKNGYQDSLTFTSPEVARNVAKPLVSYIDKAL-----VTD 298  
 294 PTS---VLFTAGHDTNLANLGALELNTLPGQDNTPTPGGELVFERWRRLSDNSQW 349  
 299 RTSAPKITVVLGHDNSIASLLTALDPKYLQHDQNTERTPIGGKIVFORWEDSKANRDLMK 358  
 350 VSLVFQTLQQRDKTPLSLNTPPGEVKLTLAGCEERNAQMGCSLAGTQIVNEA 403  
 359 IEVYQSAEQLRNADALTLOAPQRTVLELSSG-PIDADGFCPMDKPFDSVLNFA 411

RESULT 6  
 B95636  
 periplasmic glucose-1-phosphatase [imported] - Escherichia coli (strain O157:H7, subse  
 C:Species: Escherichia coli  
 C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
 C:Accession: B95636  
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayh  
 iller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimaienta, E.; Potamouisis, K.; Apodac  
 Nature 409, 529-533, 2001  
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A:Reference number: A85480; MUID:21074935; PMID:11206551  
 A:Accession: B95636  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-413 <STO>  
 A:Cross-references: GB:AE005174; NID:gl2514273; PIDN:AA55550.1; GSPDB:GN00145; UWGP:Z  
 A:Experimental source: strain O157:H7, substrain EDL933  
 C:Genetics:  
 A:Gene: agp

Query Match 25.6%; Score 551; DB 2; Length 413;  
 Best Local Similarity 33.3%; Pred. No. 1e-36;  
 Matches 138; Conservative 71; Mismatches 169; Indels 36; Gaps 12;  
 1 QSEPE-LKLESVIVSRHGVRAPTKATQMLMDVTPDAWPTWPKLGELTPRGGLIAYL 58  
 23 QTVPEGYQLQCVLMMSRHLNRLAPLANNVSVLSQSTPNKWPEDVPGGQLTTKGGVLEVM 82  
 59 GHYWRQLVADGLPKKCPQSGQVIAIADVDERTKGTGEAFAGLAPDCAITVHTQADT 118

Db 83 GHYREWLAEQGMVSGCEPPEDTVYAVANSIQRVATAQFFITGAFFGCDIPVHQBKM 142  
QY 119 SSPDLFNLKGTGVCOLDNANVTD---AILERAGSGIADFTGHYQTAFARELERVLPFQOS 175  
Db 143 GTMDPTFNPVITD---DSAAFSEQAAAMEKLSKL-QLTDSYQL-----LEKIVNYKDS 193  
QY 176 NLCKREKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLQQAQGM-EPG 233  
Db 194 PAC-----KEKQCSLVGKNTFSKYOQEGVSGPLKVGNSLVDAFTLYIEGPPMDQVA 249  
QY 234 WGRITDHWNTLLSHNAQFDLLQRTPEVARSATPPLDLIKTALTTPHPKQAYGVTL 293  
Db 250 WGEIKSDQWKVLSKLNKGYSQDSLFTSPEVARNVAKPLVSYIDKAL-----VTD 298  
QY 294 PTS---VLFAGHTNLANLGGALELN-WTLPGQPDNTPPGGELVFERWRRLSDNSQWIO 349  
Db 299 RTSAPKITVLVGHDSNIASLLTALDFKPYQLHDQNETPIGKIVFORWHDGKANRDLMK 358  
QY 350 VSLVFQTLQOMRDKTPLSLNTPPGEVKLTLAGCEERNAQMGCSLAGFTQIVNEA 403  
Db 359 IEVYQSAEQLRNADALTLOAPQQRVTLELSCG-PIDADGFCPMDKFDVSLNEA 411

## RESULT 7

JV0087  
C:Species: Escherichia coli  
C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 01-Mar-2002  
C:Accession: J00087; H6481  
C:Fraser, E.; Marck, C.; Boquet, P.L.  
J. Bacteriol. 172, 802-807, 1990  
A:Title: Nucleotide sequence and transcriptional analysis of the Escherichia coli agp gene  
A:Reference number: JV0087; MUID:90130318; PMID:2153660  
A:Accession: JV0087  
A:Molecule type: DNA  
A:Residues: 1-413 <PRA>  
A:Cross-references: GS:M33807; NID:g145217; PIDN:AAA23426.1; PID:g145218  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, J.A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: H6481  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-413 <BLAT>  
A:Cross-references: GB:AE000202; GB:U00096; NID:g1787233; PIDN:AAC74087.1; PID:g1787237;  
A:Experimental source: strain K-12, substrain MG1655  
C:Genetics:  
A:Gene: agp  
A:Map position: 23 min  
C:Function:  
A:Note: essential for growth in a high-phosphate medium containing glucose-1-phosphate  
C:Keywords: homodimer; periplasmic space; phosphohistidine; phosphoprotein; phosphoric acid  
F:1-22/Domain: signal sequence #status predicted <SIG>  
F:23-413/Product: glucose-1-phosphatase #status predicted <MAT>  
F:40/Active site: His (phosphohistidine intermediate) #status predicted  
F:311/Active site: His #status predicted

Query Match 25.3%; Score 544; DB 2; Length 413;

Best Local Similarity 33.1%; Pred. No. 3.7e-36;

Matches 137; Conservative 71; Mismatches 170; Indels 36; Gaps 12;

QY 1 QSEPE-LKLESVIVSRHGVRAP-TKATQLMODVTPDAMPWPVKLGELTPRGGLIAYL 58  
Db 23 QTPPEGYQLQVLMMSRNLRLAPLANNVLAQSTPNAPWPDVPGGLITTKGVLEVYM 82  
QY 59 GHYWRQLVADGLLEKPCGQSGQVAIIADVDERTRKTGEAPAGLAPDCAITVHTQADT 118  
Db 83 GHYREWLAEQGMVSGCEPPEDTVYAVANSIQRVATAQFFITGAFFGCDIPVHQBKM 142

QY 119 SSPDLFNLKGTGVCOLDNANVTD---AILERAGSGIADFTGHYQTAFARELERVLPFQOS 175  
Db 143 GTMDPTFNPVITD---DSAAFSEQAAAMEKLSKL-QLTDSYQL-----LEKIVNYKDS 193  
QY 176 NLCKREKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLQQAQGM-EPG 233  
Db 194 PAC-----KEKQCSLVGKNTFSKYOQEGVSGPLKVGNSLVDAFTLYIEGPPMDQVA 249  
QY 234 WGRITDHWNTLLSHNAQFDLLQRTPEVARSATPPLDLIKTALTTPHPKQAYGVTL 293  
Db 250 WGEIKSDQWKVLSKLNKGYSQDSLFTSPEVARNVAKPLVSYIDKAL-----VTD 298  
QY 294 PTS---VLFAGHTNLANLGGALELN-WTLPGQPDNTPPGGELVFERWRRLSDNSQWIO 349  
Db 299 RTSAPKITVLVGHDSNIASLLTALDFKPYQLHDQNETPIGKIVFORWHDGKANRDLMK 358  
QY 350 VSLVFQTLQOMRDKTPLSLNTPPGEVKLTLAGCEERNAQMGCSLAGFTQIVNEA 403  
Db 359 IEVYQSAEQLRNADALTLOAPQQRVTLELSCG-PIDADGFCPMDKFDVSLNEA 411

## RESULT 8

AG0632  
C:Species: Salmonella enterica subsp. enterica serovar Typhi  
A:Note: this species has also been called Salmonella typhi  
C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C:Accession: AG0632  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Fairhead, S.; Mouton, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A:Authors: Park, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.  
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi  
A:Reference number: AB0502; MUID:21534947; PMID:11677608  
A:Accession: AG0632  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-413 <PAR>  
A:Cross-references: GS:AL513382; PIDN:CAD08242.1; PID:g16502289; GSPDB:GN00176  
C:Genetics:  
A:Gene: STY1153

Query Match 25.2%; Score 543; DB 2; Length 413;

Best Local Similarity 33.6%; Pred. No. 4.5e-36;

Matches 137; Conservative 66; Mismatches 181; Indels 24; Gaps 10;

QY 1 QSEPE-LKLESVIVSRHGVRAP-TKATQLMODVTPDAMPWPVKLGELTPRGGLIAYL 58  
Db 23 QTPPEGYQLQVLMMSRNLRLAPLANNVLAQSTPNAPWPDVPGGLITTKGVLEVYM 82  
QY 59 GHYWRQLVADGLLEKPCGQSGQVAIIADVDERTRKTGEAPAGLAPDCAITVHTQADT 118  
Db 83 GHYREWLAEQGMVSGCEPPEDTVYAVANSIQRVATAQFFITGAFFGCDIPVHQBKM 142  
QY 119 SSPDLFNLKGTGVCOLDNANVTDAILERAGSGIADFTGHYQTAFARELERVLPFQOSNLC 178  
Db 143 GTMDPTFNPVITDSSAAFQQAQVA-MEKARSQ-----HLDSEYKLEIQITHYQDSPEC 196  
QY 179 LKREKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLQQAQGM-EPGWR 236  
Db 197 -----KEKQCSLIDAKTFSANYQBPQVQGLKVGNSLVDAFTLYIEGPPMDQVWGG 252  
QY 237 ITDHWNTLLSHNAQFDLLQRTPEVARSATPPLDLIKTALTTPHPKQAYGVTLPTS 296  
Db 253 IHTDQWKVLSKLNKGYSQDSLFTSPTVARNVAKPLVSYIDKVLV-----AERVSAP-K 304  
QY 297 VLFAGHTNLANLGGALELN-WTLPGQPDNTPPGGELVFERWRRLSDNSQWIOVSLVFQ 355  
Db 305 VTLVGHDSNIASLLTALDFKPYQLHDQNETPIGGLVFORWHDGKANRDLMKTYVQ 364  
QY 356 TLOOMRDKTPLSLNTPPGEVKLTLAGCEERNAQMGCSLAGFTQIVNEA 403

Db 365 SARQLRNAEALTLKSPQAVTLELKGCPV-DANGFCPLDKFDNMNTA 411  
RESULT 9  
E87316  
periplasmic phosphoanhydride phosphohydrolase [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C>Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C:Accession: E87316  
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete Genome Sequence of *Caulobacter crescentus*.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: E87316  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-414 <STO>  
A:Cross-references: GB:AF005673; NID:g13421731; PIDN:AAK22529.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CC0542

Query Match 24.8%; Score 534.5; DB 2; Length 414;  
Best Local Similarity 33.7%; Pred. No. 2.2e-35;  
Matches 137; Conservative 60; Mismatches 177; Indels 33; Gaps 9;  
QY 8 LESVIVSRHGVRAPTKATQMLQMDVTPDAMPVVKLGELTPRGGELIAYLGHYRQRLV 67  
Db 25 LEKVILSRHGVRASMSPERLEASARPWRFEPVPAHULTARGETLVASMGDYRHYA 84  
QY 68 ADGLLPKCGCQSQVAIIADVDRTRKTEGAPAGLAPDCAITVHTQADTSPDPLFNP 127  
Db 85 AQGLLPKCGDC---ASVYAWANVTORTATKAYRETLAPGCPVTNTVGE-GNIDPMFEP 140  
QY 128 LKTVGCOLDNANVTDAILERAGGSIADFTGHYQTAFARELRVNFQSNCLKREKQDES 187  
Db 141 VKAGIVKADHALARAANVAGRVGGDLTANASHNQEAQLDALLMQCDKGPC----- 191  
QY 188 CSLTQALPSELUK-----SADCVSLTGAVSLASMLTEIFLLQQAQMPGCG-RI 237  
Db 192 ----PPAPGKRVFPAKFGFYDGBELAGLSGPEAFASGVTESLMAWADGRDFAGLGWKS 247  
QY 238 TDSHQWNTLLSHNAQFDLQRTPEVARSRAATPLDLIKTALTTPHPKQKQAYVTLTSPV 297  
Db 248 LDEALTRSFLLHQAEPLRLRTTPVARTLAGHLADRLAATLR---DGAAAIG-PVDARL 303  
QY 298 LFIAGHDTNLANLGGALELNLTPG-QPDNTPPGGELVFERWRRLSDNSQWISLVFQT 356  
Db 304 VTIAGHDGTLASGLLRLMWTLPQYQNPQIQPGGALVFERWR-DDGVRVVRVRFQGS 362  
QY 357 LQMRDKTPLSLNTPPGGVKTLACBERRNAQGCSLAGFTQIVNEA 403  
Db 363 LSQRNMTALDAKTPPLSAPVFGCGGTATPAFDCRLDFETVVRGA 409

RESULT 10  
S25627  
Glucose-1-phosphatase (EC 3.1.3.10) precursor - Providencia rettgeri  
C:Species: Providencia rettgeri  
C>Date: 25-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 08-Oct-1999  
C:Accession: S25627  
R:Riccio, M.L.; Chiesurin, A.; Lombardi, G.; Satta, G.  
submitted to the EMBL Data Library, September 1992  
A:Reference number: S25627  
A:Accession: S25627  
A:Molecule type: DNA  
A:Residues: 1-417 <RIC>  
A:Cross-references: EMBL:X68201; NID:g45771; PIDN:CAA48288.1; PID:g45772  
A:Experimental source: strain PV7  
C:Genetics:  
A:Gene: ggp  
C:Keywords: periplasmic space; phosphohistidine; phosphoprotein; phosphoric monoester by

F1-29/Domain: signal sequence #status predicted <SIG>  
F130-417/Product: glucose-1-phosphatase #status predicted <MAT>  
F142/Active site: His (phosphohistidine intermediate) #status predicted  
F1314/Active site: His #status predicted

Query Match 24.1%; Score 518; DB 2; Length 417;  
Best Local Similarity 30.4%; Pred. No. 4.7e-34;  
Matches 126; Conservative 83; Mismatches 167; Indels 38; Gaps 11;  
QY 2 SEPELKLSVIVSRHGVRAPTKATQMLQMDVTPDAMPVVKLGELTPRGGELIAYLGHY 61  
Db 27 NQADWLVQVLSLHNRRTIVNTGILTEVTDKKWPMDAKSGVLTITQGGALEVYMGHY 86  
QY 62 WRQ-----RLVADGLLPKCGCQSQ-QVAIIADVDRTRKTEGAPAGLAPDCAITVHTQ 115  
Db 87 FREWIDQNKLADEL-----CPTSNEDIYLYTNSLQRTIATAQFPAAQAFPGCKVNIHQ 141  
QY 116 ADTSSPDPLENPLKTVGCOLDNANVTDAILERAGGSIADFTG---HYQTAFARELRVNLFP 173  
Db 142 PEIGKNDPVFNPIIT-----NGSPFKQKALAMDDYLKGLSLKAGYBELDTVLNIK 193  
QY 174 QSNCLKREKQDESCLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMP--E 231  
Db 194 DSOKC---KTDKLCNLDQKNSFIIEADKEPVGSGPLKIANSAVDAIDLQYEGFPADQ 249  
QY 232 PFGWRTDTSQWNTLLSHNAQFDLQRTPEVARSRAATPLDLIKTALTTPHPKQKQAYGV 291  
Db 250 VAWGLVDTPFERKWLKLNAYQETLTPKTIKKNVAPILNYIDKGVFVVDKGETA--- 306  
QY 292 TLPTSVLFTAGHDTNLANLGGALELN-WLTPGQPDNTPPGGELVFERWRRLSDNSQWITQV 350  
Db 307 ----KFILVGHDSNIASIMSAMDPKYQLAQOYEHTPIGGKLVFORWTDQTKKDFMKV 362  
QY 351 SILVFQTLQMRDKTPLSLNTPPGGVKTLACGE-ERNAQGCSLAGFTQIVNEA 403  
Db 363 EYVYQTADQLRDNAYLSLETPEKPVHTLELKCQPVDKN--GYCSWEDFOKWMKA 414

RESULT 11  
T16058  
Hypothetical protein F13D11.1 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 23-Mar-2001  
C:Accession: T16058  
R:Fulton, L.  
submitted to the EMBL Data Library, November 1995  
A:Description: The sequence of *C. elegans* cosmid F13D11.  
A:Reference number: S69020  
A:Accession: T16058  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-416 <FUL>  
A:Cross-references: EMBL:U40939; NID:g1072175; PID:g1072178; PIDN:AAA81702.1; CESP:F131  
C:Genetics:  
A:Gene: CESP:F13D11.1  
A:Introns: 23/3; 59/3; 88/3; 138/3; 202/3; 244/1; 270/3; 312/3; 349/3; 377/1  
C:Superfamily: mammalian acid phosphatase

Query Match 6.2%; Score 133; DB 2; Length 416;  
Best Local Similarity 21.5%; Pred. No. 0.0063;  
Matches 92; Conservative 57; Mismatches 161; Indels 118; Gaps 18;  
QY 7 KLSVIVSRHGVRAPTKATQMLQMDVTPD-----AWPTVPVKLGELTPRGGELIAYLGH 60  
Db 17 KLFVQVIFRFGERTPGHYLYFPGDDLNNVYQQTAWP-----GLTKRGILERFQLG- 69  
QY 61 YMRQL-----VADGLLPKCGCQSQGVIAIADVDRTRKTEGAPAGLAPDCAITVH 113  
Db 70 ---QRLKRYGEFGTYQPR-----DFHYVTGKDNRTSASQAQAFAGFLPP----- 113  
QY 114 TQADTSPDPLFNPLKTVGCOLDNANVTDAILERAGGSIADFTGHYQTAFARE-----L 166  
Db 114 NEDQTNWYELKQPV-----AQLTDESIDWSLGAIDNCPVYGEAQRKSSSEYAEVM 164

```

167  ERVLNPPOSNLCRKEEKODESCSLTQALPSELKVSADCVSLTGAVSLASMLTBI---FLL 223
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
165  DOMEXYDAELLQVLVENHADE-----PIVSAVKYNHVIDSLSKRYIL 205
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
224  QQAQMPFPGWGRITDTSHQWNTLASLHNAQFDLLQRTPEVARSRATEFLD---LIKTL 279
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
206  QQDR-LPYEPWARGYENRILNMSFLIHDA-----VVKQNSDVGVDHNELVMSYF 254
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
280  TPHPQKQAYGTYLPTSVLFTAGHDTNLNIAAGALELNWTLPGOPNTTP-GGELVFERW 338
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
255  ETHLQKNSKGV-----FISGHDNLVTTWESURLD---GHPEDIPNYGHAHIAE--- 301
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
339  RRLSDNSQWIVSLVFQTLQWRDKYTPLSLNTPGCEVNLTLACGEERNAGCMGSLAGTQQ 398
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
302  -----MHEPVGQLSIKFFFLMSGFNQTRVELPHFPC-SKSQNNDCDWDDEFOR 346
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
399  IVNEARIP 406
    : : : : :
347  LVKKSRRKP 354
    : : : : :

```

RESULT 12

S06167  
acid phosphatase (EC 3.1.3.2) ACP2 precursor [validated] - human  
N;Alternate names: acid phosphatase, lysosomal  
C;Species: Homo sapiens (man)  
C;Date: 04-Dec-1992 #sequence revision 01-Dec-1995 #text\_change 08-Dec-2000  
C;Accession: S06167; S05525; S01155  
R;von Figura, K.  
submitted to the EMBL Data Library, June 1989  
A;Reference number: S06167  
A;Accession: S06167  
A;Molecule type: DNA  
A;Residues: 1-423 <ON>  
A;Cross-references: EMBL:X15525; NID:G934239; PIDN:CAA33542.1; PID:G1199524  
R;Geier, C.; von Figura, K.; Pohlmann, R.  
Eur. J. Biochem. 183, 611-616, 1989  
A;Title: Structure of the human lysosomal acid phosphatase gene.  
A;Reference number: S05525; MUID:89377828; PMID:2776754  
A;Accession: S05525  
A;Molecule type: DNA  
A;Residues: 1-29 <GI>  
R;Pohlmann, R.; Krentler, C.; Schmidt, B.; Schroeder, W.; Lorkowski, G.; Cullen  
EMBO J. 7, 2343-2350, 1988  
A;Title: Human lysosomal acid phosphatase: cloning, expression and chromosomal  
A;Reference number: S01155; MUID:89052645; PMID:3191910  
A;Accession: S01155  
A;Molecule type: mRNA  
A;Residues: 1-423 <POH>  
A;Cross-references: EMBL:X12548; NID:G34262; PIDN:CAA31064.1; PID:G34263  
A;Note: part of this sequence, including the amino end of the mature protein,  
C;Genetics:  
A;Gene: GDB:ACP2  
A;Cross-references: GDB:118963; OMIM:171650  
A;Map position: 11p11.2-11p11.1  
A;Introns: 38/3; 70/3; 99/3; 150/3; 183/3; 213/3; 285/3; 321/2; 380/1  
C;Function:  
A;Description: catalyzes the hydrolysis of a wide range of phosphate esters  
C;Superfamily: mammalian acid phosphatase  
C;Keywords: glycoprotein; phosphatidine; phosphoprotein; phosphoric monoester  
F;1-30/Domain: signal sequence #status predicted <SIG>  
F;31-423/Product: acid phosphatase ACP2 #status experimental <MAT>  
F;41/Active site: Arg #status predicted  
F;42/Active site: His (phosphatidine intermediate) #status predicted  
F;92,133,167,171,267,322,331/Binding site: carbohydrate (Asn) (covalent)  
F;159-370,212-310,345-349/Disulfide bonds: #status predicted

Query Match 6.1%; Score 132; DB 1; Length 423;  
Best Local Similarity 23.3%; Pred. No. 0.0077;  
Matches 97; Conservative 43; Mismatches 173; Indels 104; Gaps 20

Db	33	LFVTLVYRHGDRSPVKT----	YPKDPYQBEWPQFGQLTKGMLQHWELGQALRQRY-	87
Qy	68	ADGLLPKCGCPQSGOVAIIADV-	ERRTKGTGEAFAAGLAPDCAITVHTQADTSSPDDL--	124
Db	88	-HGFL-----NTSVHRQEVVVRSTDFR	TLMSAEANLAGLFP-----PNGMQR	129
Qy	125	FNP-LKTGVCGOLDNANVTDA	ILERAGGSIADFTGHYQTAPRELERVLNFPQSNCLKREK	183
Db	130	FNFNISWQPIPVHTVPIITE-----	DRLLKFLPGCPRYEQL	165
Qy	184	QDESCLTQALPSELKVSADCV	SLTGAVSLASMLTEIFL-----LQAQGMPEP	232
Db	166	QNE-----TRQTPQVQNSSRNAQ	FLDMVANETGLDTLTETVNVYDTLFCQTHGLRLP	221
Qy	233	GMGRITDQHWNTLLSHNAQFDLL--	QRTPEVARSRAATPLDLTKTALTTPHPKQAY	289
Db	222	PW---ASPQTQWRLSRLDQF	SFREFLFGIYQQAERLQGGVLLAQIRKNLTLMATTSQ--	276
Qy	290	GVTLPSTVLFTAGHDTNLAN	LGGALELNWTLPGQPDNTPPGSELVFERWRRLSDNSQWIQ	349
Db	277	---LP-KULVTSADHTTVL	VALQMALDVI-----NGEQAPYASCHIFELQEDSEN--	PS 324
Qy	350	VSLVFQTLQQMRDKTPLSLNT	PPGSEVKLTLAGCEERNAQMGCSLAGTQIVNEARIP	406
Db	325	VEMYF---RNESDKAPWLSLP-	-----GCPHR-----CPLQDFLRL-TEPVVP	363
RESULT 13				
A33395				
acid phosphatase (EC 3.1.3.2) precursor - rat				
C;Species: Rattus norvegicus (Norway rat)				
C;Date: 29-Jan-1990 #sequence_revision 29-Jan-1990 #text_change 22-Jun-1999				
C;Accession: A33395				
R;Himeno, M.; Fujita, H.; Noguchi, Y.; Kono, A.; Kato, K.				
Biochem. Biophys. Res. Commun. 162, 1044-1053, 1989				
A;Title: Isolation and sequencing of a cDNA clone encoding acid phosphatase in				
A;Reference number: A33395; MUID:89350910; PMID:2764916				
A;Accession: A33395				
A;Status: preliminary				
A;Molecule type: mRNA				
A;Residues: 1-423 xHM-				
A;Cross-references: GB:M27893; NID:G202933; PIDN:AAA0744.1; PID:G202934				
C;Superfamily: mammalian acid phosphatase				
C;Keywords: phosphohistidine; phosphoprotein; phosphoric monoester hydrolase				
F;41/Active site: Arg #status Predicted				
F;42/Active site: His (phosphohistidine intermediate) #status Predicted				
Query Match 5.3%; Score 114; DB 2; Length 423;				
Best Local Similarity 22.7%; Pred. No. 0.22; Indels 106; Gaps 20;				
Matches 95; Conservative 42; Mismatches 175; Indels 106; Gaps 20;				
Qy	6	LESVIVSRHGVRAPTKATQLM	QMDVTPDPAWTPVVKLGELTPRGGELIAYLGHYWRQLV	67
Db	33	LFVTLVYRHGDRSPVKA----	YPKDPYQBEKWPQFGQLTKGMLQHWELGQALRQRY-	87
Qy	68	ADGLLPKCGCPQSGOVAIIADV-	ERRTKGTGEAFAAGLAPDCAITVHTQADTSSPDDL	126
Db	88	-HGFL-----NASYHREVVVVRSTDF	FRTLMSAEANLAGLFP-----TEVCHFNPNISWQ	137
Qy	127	PLKTVCGOLDNANVTDA	ILERAGGSIADFTGHYQTAPRELERVLNFPQSNCLKREKQDE	186
Db	138	PI-----PVHTVPIITE-----	DRLLKFLPGCPRYEQLONE	168
Qy	187	PSCLTQALPSELKVSADCV	SLTGAVSLASMLTEIFL-----LQAQGMPEP	235
Db	169	---TRQTPQVQNSSRNAQ	FLDMVANETGLMNLTEIWNVYDTLFCQTHGLLLPPW-	223
Qy	236	RITDQHWNTLLSHNAQF-----	DLQSTPEVARSRAATPLDLTKTALTTPHPKQOA	288
Db	224	--ASPQTQALSQLDFSL	FLFGIHQVQK-----ASLQGVLLAQILKNLTLMATTSQF	277
Qy	289	YGVTLPTSVLFTAGHDTNL	ANLGGALELNWTLPGQPDNTPPGSELVFERWRRLSDNSQWI	348



278 -----PKLLVSAHDTTVALQALNVY-----NGKQAPVASCHIFELVQ--EDNGNF- 323  
QY 349 QVSLVFQTLQMRDKTSLNTPGCVKLTLAGCEERNAQCMCSLAGTQIVNEARIP 406  
Db 324 SVEMVFRNDSK---KAPW-----PLTLPGCPHR-----CPLQDFLKL-TEFVIP 363

RESULT 14  
S64682  
acid phosphatase (EC 3.1.3.2) AcpH-1 precursor, lysosomal - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C:Date: 06-Dec-1996 #sequence\_revision 06-Dec-1996 #text\_change 17-Mar-2000  
C:Accession: S64682; S64681  
R:Chung, H.J.; Shaffer, C.; MacIntyre, R.  
Mol. Gen. Genet. 250, 635-646, 1996  
A:Title: Molecular characterization of the lysosomal acid phosphatase from Drosophila melanogaster  
A:Reference number: S64681; MUID:96194627; PMID:8676866  
A:Accession: S64682  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-438 <CHU>  
A:Accession: S64681  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-31, 'Y', 33-438 <CHW>  
C:Genetics:  
A:Gene: FlyBase:AcpH-1  
A:Cross-references: FlyBase:FBgn0000032  
C:Superfamily: mammalian acid phosphatase  
C:Keywords: phosphoric monoester hydrolase  
F:1-33/Domain: signal sequence #status predicted <SIG>  
F:34-438/Product: acid phosphatase AcpH-1 #status predicted <MAT>

Query Match 5.2%; Score 112.5; DB 2; Length 438;  
Best Local Similarity 20.3%; Pred. No. 0.31;  
Matches 82; Conservative 59; Mismatches 145; Indels 117; Gaps 19;

QY 7 KLESVIVSRHGVRAPTKATQLMQDVTTPDAWPT-----WPVKLGELTPRGGLTAYL 58  
Db 47 QLKFWVHYRGDRTPV-----DPPTDFWGRKFWPTGWDLTNLGKQEHYDL 95  
QY 59 GHYWRQRLVADGLPKKCCPOSGQVIAIDVD--ERTRKTGEAFAGL-----APDC----- 108  
Db 96 GKMLRNRY--SNLLPPI---YSNENIYVQSTDVDRTLMSAQSNLAGLYEPQGEDIWNTDI 150  
QY 109 ---AITVHTQADTSSPDPLNPLKTVGCOLDNANVTDAILERAGGSIADTGHYQTAFARE 165  
Db 151 NWQPIPIHTSPERDDP-----ILAAKAPCPAYDY-----E 180  
QY 166 LERLVNFPQSNLCLKREKQDESCSLTQALPSELKVSADCVSLTGAVALSLMLTEIFLQQ 225  
Db 181 LASLESSPEFK-ALTEKRNLFAYLSEKGGRPVKTFFD-----AQYLANLTFIEN 229  
QY 226 AQGWPEPCWGRITDISHQWNTLLSLHNAQFDLLQRTPEVARSRATPLL-DLIKALTALTPHP 284  
Db 230 LYNNLTLPKWTCKVYGRELATYVS--NFAFAISSYTRKLARKAGFLKIDIFORP----- 281  
QY 285 QKQAYGVTLPTSVLFI-AGHDNTLANLGALLENWTLPGQPDNTPPGGELVFERWRLSD 343  
Db 282 KEKSSGSLKPDSSMWVYSAHDTTVAVLNALKL-FEL-----HSPP----- 321  
QY 344 NSQWIQVSLVFQTLQMRDKTPL-----SLNTPPGCVKLTLAGC 382  
Db 322 -----YTACIMMELRVDETNPFLVSIFYKNTTASPLPLDIPGC 359

RESULT 15  
B89130  
protein F52E1.8 [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 03-Mar-2003  
C:Accession: B89130

R:anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology  
A:Reference number: A75000; MUID:99069613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_e-  
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; a  
A:Accession: B89130  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-344 <STO>  
A:Cross-references: GB:chr\_V; PIDN:AB37044.1; PID:gl086806; GSPDB:GN00023; CESP:F52E1  
C:Genetics:  
A:Gene: F52E1.8  
A:Map position: 5  
C:Superfamily: mammalian acid phosphatase

Query Match 5.1%; Score 109.5; DB 2; Length 344;  
Best Local Similarity 19.2%; Pred. No. 0.38;  
Matches 82; Conservative 63; Mismatches 138; Indels 145; Gaps 19;

QY 16 RHGVRAPTKATQLMQDVTTPDAWPT-WPVKLGELTPRGGLTAYLGHYWRQRLVADGLLPK 74  
Db 6 RHGDRAPGDLPPKDKYNETFWPGW----DQLTKNGIWAQVAVELGWLGRQRYGAT-VLP- 59  
QY 75 CCGPQSGQVIAIDVDERTKTRTGEAFAGLAP--DCAITVHTQADTSSPDPLNPLKTVG 132  
Db 60 --IFDKQKVFILSSDSERAIEAQAQVSAGLFPVDDRVWESSYLRYWQTPI----- 109  
QY 133 COLDNANVTDAILERAGGSIADTGHYQTAFARELSRVL-----NFPQSNLCLKREKQ 184  
Db 110 -----QTAYCTIDALLRPTKVKCFNYDLAN-----EQ 136  
QY 185 DESCSLTQALPSELKVSADCVSLTGAVALSLM-LTEIFLLOA--QGWPEPCW----- 234  
Db 137 EAPATQINNEYGQVFNWLNQITGMESIDFWNINDYDIQRELDHNMPPQWLNQVFN 196  
QY 235 -----GRITDSHOWNTLLSLHNAQF-----DLLQRTPEVARSRATPLDLIKT 277  
Db 197 TTIMDHIRELKEITRQEFN---SPTKAFRGGLVNGFQWLMEDLKANKTT----- 245  
QY 278 ALTPHPPOKQAYGVTLPTSVLFIAGHDNTLANLGALLENWTLPGQPDNTPPGGELVFER 337  
Db 246 -----KNAMYSHSDGTLTSLALYALNVS-----NDQLVPYATATVLEL 283  
QY 338 WRRLSDNSQWIOVSLVFQTLQMRDKTSLNTPPGCVKLTLAGCEERNAQCMCSLAGT 397  
Db 284 Y---DNT-----VQLFYK-----NTTSTAYPMTIPGCGQ-----ICPYSQL 318  
QY 398 QIVNEARI 405  
Db 319 QLENVRV 326

Search completed: April 30, 2004, 12:42:20  
Job time : 17 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model  
Run on: April 30, 2004, 12:31:10 ; Search time 10 Seconds  
(without alignments)  
2134.876 Million cell updates/sec

Title: US-10-021-723B-13  
Perfect score: 2153  
Sequence: 1 QSBPELKESVIVSRHGVR.....CSLAGFTQIVNEARIPACSL 410

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2077	96.5	432	1 PPA_ECOLI	P07102 escherichia
2	548	25.5	413	1 AGP_SALTY	O33921 salmonella
3	544	25.3	413	1 AGP_ECOLI	P19926 escherichia
4	518	24.1	417	1 AGP_PROSE	O52309 providencia
5	132	6.1	423	1 PPAI_HUMAN	P11117 homo sapien
6	114	5.3	423	1 PPAI_RAT	P20611 rattus norv
7	111.5	5.2	457	1 PAX8_MOUSE	Q00288 mus musculu
8	109.5	5.1	479	1 PAX8_MOUSE	P34755 aspergillus
9	109	5.1	5127	1 RY44_DROME	P47240 canis famil
10	108.5	5.0	479	1 PAX8_MOUSE	P47240 canis famil
11	107.5	5.0	459	1 PAX8_MOUSE	P47240 canis famil
12	105	4.9	450	1 PAX8_MOUSE	P47240 canis famil
13	104	4.8	423	1 PPAI_MOUSE	P24638 mus musculu
14	103	4.8	769	1 DIVL_CAUCR	Q97999 caulobacter
15	103	4.8	1048	1 SBCC_ECOLI	P13458 escherichia
16	101.5	4.7	386	1 PPAI_HUMAN	P15309 homo sapien
17	100	4.6	635	1 TPOR_HUMAN	P40238 homo sapien
18	95.5	4.4	419	1 ARG3_SVNV3	P74122 s arginine
19	94	4.4	341	1 TIN2_MOUSE	Q98299 mus musculu
20	94	4.4	612	1 MCR_XENLA	O18738 xenopus lae
21	94	4.4	895	1 DAGI_BOVIN	P20646 rattus norv
22	93.5	4.3	381	1 PPAI_RAT	P20646 rattus norv
23	92.5	4.3	625	1 TPOR_MOUSE	Q08351 mus musculu
24	92.5	4.3	702	1 CANB_HUMAN	Q9umq6 homo sapien
25	92.5	4.3	1085	1 CASR_BOVIN	P35384 bos taurus
26	91	4.2	437	1 LHA2_HUMAN	O00534 homo sapien
27	91	4.2	458	1 PAX8_RAT	P51974 rattus norv
28	91	4.2	551	1 CAP_SCHPO	P36621 schizosacch
29	91	4.2	3033	1 POLG_HCVJ6	P26650 h genome po
30	90	4.2	328	1 PAX1_PSEAE	Q915u4 pseudomonas
31	90	4.2	461	1 ATOC_ECOLI	Q06065 escherichia
32	90	4.2	921	1 T2D3_DROME	P47825 drosophila
33	90	4.2	1079	1 CASR_MOUSE	Q9qy96 mus musculu

34	89.5	4.2	369	1 HEM3_PEA	Q43082 piseum sativ
35	89.5	4.2	406	1 GSPF_VIBCH	P45780 vibrio chol
36	89.5	4.2	972	1 ORC4_SCHPO	Q9V794 schizosacch
37	89.5	4.2	1078	1 CASR_HUMAN	P41180 homo sapien
38	89	4.1	448	1 TBE1_ECHMU	Q9af27 echinococcu
39	89	4.1	811	1 PRIA_RHURU	P04445 rhodospirill
40	89	4.1	1464	1 NCO2_HUMAN	Q15596 homo sapien
41	88.5	4.1	475	1 MUKF_HAEIN	P45185 haemophilus
42	88.5	4.1	872	1 IP3L_RAT	P42335 rattus norv
43	88.5	4.1	878	1 CAPP_YERPE	Q8za84 yersinia pe
44	88.5	4.1	1531	1 NTF5_HUMAN	O94916 homo sapien
45	88.5	4.1	2554	1 7LBS_DROME	P13368 drosophila

ALIGNMENTS

RESULT 1	PPA_ECOLI	STANDARD;	PRT;	432 AA.
ID	PPA_ECOLI			
AC	P07102;			
DT	01-APR-1988 (Rel. 07, Created)			
DT	01-AUG-1991 (Rel. 19, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Periplasmic appa protein precursor [Includes: Phosphoanhydride			
DE	phosphohydrolase (EC 3.1.3.2) (pH 2.5 acid phosphatase) (AP); 4-			
DE	phosphatase (EC 3.1.3.26)]			
GN	APPA OR B0980.			
OS	Escherichia coli.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Escherichia.			
OX	NCBI_TaxID=562;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 23-29.			
RC	STRAIN=K12;			
RX	MEDLINE=90368616; PubMed=2168385;			
RA	Dassa J., Marck C., Boquet P.L.;			
RT	"The complete nucleotide sequence of the Escherichia coli gene appA			
RT	reveals significant homology between pH 2.5 acid phosphatase and			
RT	glucose-1-phosphatase."			
RL	J. Bacteriol. 172:5497-5500 (1990).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=K12 / MG1685;			
RX	MEDLINE=97426617; PubMed=9278503;			
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,			
RT	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,			
RT	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,			
RT	Mau B., Shao Y.;			
RT	"The complete genome sequence of Escherichia coli K-12."			
RL	Science 277:1453-1474 (1997).			
RL	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=K12;			
RX	MEDLINE=97061202; PubMed=8905232;			
RA	Oshima T., Aliba H., Baba T., Fujita K., Hayashi K., Honjo A.,			
RT	Ikeno K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,			
RT	Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,			
RT	Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,			
RT	Sanpei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,			
RT	Yano M., Horiuchi T.;			
RT	"A 718-Kb DNA sequence of the Escherichia coli K-12 genome			
RT	corresponding to the 12.7-28.0 min region on the linkage map."			
RL	DNA Res. 3:137-155 (1996).			
RL	[4]			
RP	SEQUENCE OF 1-112 FROM N.A.			
RX	MEDLINE=87271766; PubMed=3038201;			
RT	Touati E., Danchin A.;			
RT	"The structure of the promoter and amino terminal region of the pH			
RT	2.5 acid phosphatase structural gene (appA) of E. coli: a negative			
RT	control of transcription mediated by cyclic AMP."			
RL	Biochimie 69:215-221 (1987).			
RL	[5]			

RP SEQUENCE OF 1-17 FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=92049231; PubMed=1658595;  
RA Dassa J., Fsihi H., Marck C., Dion M., Kieffer-Bontemps M.,  
RA Boquet P.L.;  
RT "A new oxygen-regulated operon in *Escherichia coli* comprises the  
RT genes for a putative third cytochrome oxidase and for pH 2.5 acid  
RT phosphatase (appa).";  
RL Mol. Gen. Genet. 229:341-352(1991).  
[6]  
RN CHARACTERIZATION, AND SEQUENCE OF 23-34.  
RA Greiner R., Jany K.D.;  
RL "Characterization of a phytase from *Escherichia coli*.";  
RN Biol. Chem. Hoppe-Seyler 372:664-665(1991).  
[7]  
RN CHARACTERIZATION, AND SEQUENCE OF 23-35.  
RX MEDLINE=93256556; PubMed=8387749;  
RA Greiner R., Konietzny U., Jany K.D.;  
RT "Purification and characterization of two phytases from *Escherichia coli*.";  
RL Arch. Biochem. Biophys. 303:107-113(1993).  
[8]  
RN CHARACTERIZATION.  
RP MEDLINE=20161462; PubMed=10696472;  
RA Golovan S., Wang G., Zhang J., Forsberg C.W.;  
RA van Etten R.L.;  
RT "Overexpression, site-directed mutagenesis, and mechanism of  
RT *Escherichia coli* acid phosphatase";  
RL J. Biol. Chem. 267:22830-22836(1992).  
[9]  
RN X-RAY CRYSTALLOGRAPHY (2.05 ANGSTROMS).  
RP MEDLINE=20122624; PubMed=10655611;  
RA Lim D., Golovan S., Forsberg C.W., Jia Z.;  
RT "Crystal structures of *Escherichia coli* phytase and its complex with  
RL phytate";  
RL Nat. Struct. Biol. 7:108-113(2000).  
CC -1- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an  
CC alcohol + phosphate.  
CC -1- CATALYTIC ACTIVITY: Myo-inositol hexakisphosphate + H(2)O = 1D-  
CC myo-inositol 1,2,3,4,5-pentakisphosphate + phosphate.  
CC -1- SUBUNIT: Monomer.  
CC -1- SUBCELLULAR LOCATION: Periplasmic.  
CC -1- INDUCTION: IN ADDITION TO CAMP-MEDIATED CONTROL, THIS ENZYME IS  
CC INDUCED WHEN BACTERIAL CULTURES REACH STATIONARY PHASE. ITS  
CC SYNTHESIS IS TRIGGERED BY PHOSPHATE STARVATION OR A SHIFT FROM  
CC AEROBIC TO ANAEROBIC CONDITIONS.  
CC -1- SIMILARITY: Belongs to the histidine acid phosphatase family.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; M58708; AAB2086.1; -;  
DR EMBL; AE000200; AAC74065.1; -;  
DR EMBL; D90735; BAA35745.1; -;  
DR EMBL; X05471; CAA29031.1; -;  
DR EMBL; M58708; -; NOT ANNOTATED CDS.  
DR EMBL; S63811; AAB20286.1; -;  
DR PIR; B36733; B36733.  
DR PDB; 1DKL; 03-AUG-00.  
DR PDB; 1DKM; 02-AUG-00.

DR PDB; 1DKN; 03-AUG-00.  
DR PDB; 1DKO; 03-AUG-00.  
DR PDB; 1DKP; 03-AUG-00.  
DR PDB; 1DKQ; 03-AUG-00.  
DR SWISS-2DPAGE; P07102; COLI.  
DR EcoGene; EGI0049; appA.  
DR InterPro; IPR000560; HisAc\_phsphtse.  
DR Pfam; PF00328; acid\_phosphat\_1; 1.  
DR PROSITE; PS00616; HIS\_ACID\_PHOSPHAT\_1.  
DR PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.  
KW Hydrolase; Periplasmic; Signal; Multifunctional enzyme; 3D-structure;  
KW Complete proteome.  
FT SIGNAL 1 22  
FT CHAIN 23 432  
FT ACT\_SITE 39 39  
FT ACT\_SITE 326 326  
FT DISULFID 99 130  
FT DISULFID 155 430  
FT DISULFID 200 410  
FT DISULFID 404 413  
FT CONFLICT 51 66  
FT CONFLICT 75 76  
FT CONFLICT 112 112  
FT STRAND 28 38  
FT STRAND 42 42  
FT HELIX 49 53  
FT TURN 54 54  
FT TURN 66 67  
FT STRAND 69 69  
FT HELIX 71 90  
FT TURN 91 92  
FT TURN 102 103  
FT STRAND 105 109  
FT HELIX 113 126  
FT TURN 128 129  
FT STRAND 134 135  
FT TURN 140 141  
FT HELIX 145 147  
FT TURN 149 153  
FT HELIX 159 169  
FT TURN 170 171  
FT HELIX 174 179  
FT TURN 180 181  
FT HELIX 182 192  
FT TURN 193 193  
FT HELIX 194 196  
FT HELIX 198 201  
FT TURN 202 202  
FT HELIX 212 215  
FT STRAND 220 222  
FT STRAND 227 229  
FT HELIX 231 248  
FT TURN 249 250  
FT HELIX 254 257  
FT TURN 258 258  
FT HELIX 262 279  
FT TURN 280 281  
FT HELIX 283 289  
FT TURN 290 290  
FT HELIX 291 301  
FT STRAND 307 308  
FT HELIX 310 312  
FT STRAND 314 315  
FT STRAND 319 324  
FT HELIX 326 336  
FT TURN 337 337  
FT STRAND 342 342  
FT TURN 343 344  
FT STRAND 345 345  
FT TURN 351 352  
FT STRAND 354 362  
FT TURN 363 366

MODVTPDAWPTWPKL -> NAGCHPRRMANLAGKT (IN  
REF. 3).  
EL -> DV (IN REF. 4).  
D -> S (IN REF. 4).

PERIPLASMIC APPA PROTEIN.  
NUCLEOPHILIC ACCEPTOR.  
PROTON DONOR.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; AE008748; AAL20049.1; --

CC EMBL; U75949; AAC45604.1; --

CC HSP; P07102; IDXL.

CC StyGene; SG10595; agp.

CC InterPro; IPR00560; HisAc\_pheptase.

CC Pfam; PF00328; acid\_phosphat; 1.

CC PROSITE; PS00616; HIS\_ACID\_PHOSPHAT\_1; 1.

CC PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.

CC Hydrolase; Periplasmic; Signal; Complete proteome.

CC SIGNAL 1 22 BY SIMILARITY.

CC FT CHAIN 23 413 GLUCOSE-1-PHOSPHATASE.

CC FT ACT SITE 40 40 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).

CC FT ACT SITE 311 311 PROTON DONOR (BY SIMILARITY).

CC FT CONFLICT 57 57 T -> P (IN REF. 2).

CC FT CONFLICT 60 63 AMPA -> TCPP (IN REF. 2).

CC FT CONFLICT 74 74 K -> T (IN REF. 2).

CC SEQUENCE 413 AA; 45559 MW; E057667C7F8A7244C CRC64;

Query Match 25.5%; Score 548; DB 1; Length 413;  
Best Local Similarity 33.8%; Pred. No. 8.1e-37;  
Matches 138; Conservative 66; Mismatches 180; Indels 24; Gaps 10;

QY 1 QSEPE-LKLESVIVSRHGVRAPIKATQMDVTPDAWTPVKLGELTPRGGLIAYL 58  
DB 23 QTTPEGYQLQVLMMSRHLRAPLANNGSVLTAQSTFNAPAWDVPGGQLTTKGGVLEVYM 82

QY 59 GHYWRQLVADGILLPKCGCPQSGQVAILADVDERTKGTGEAFAGLAPDCAITVHTQADT 118  
DB 83 GHYTRMLVAQGLIPSGECFAPDTVYVANSIQRVATVATQAFITGAPGCDIPVHHEKM 142

QY 119 SSPDPLFNPLKTVGCOLDNANVTDAILERAGSIADFTGHYQTAFAELERVLNFPQSNLC 178  
DB 143 GTMDPTFNFTVITDSDAAFRQAVQA-MEKARSOL-----HLDES YKLLSQITHYQDSPSC 196

QY 179 LKREKQDESCSLTQALPSELKVSADCVSLTGAVSLSMILTEIFLQQAQGMPEFPGWR 236  
DB 197 -----KEKHCQLIDAKDTPSANVQBPVGQPLKGVNSLVDAFTLQYEGFPMDQVAMGG 252

QY 237 ITDSHOWNTLLSHNAQFDLLQRTPEVARSRATPLDLIKTALTTPHPKQAYGTLPTS 296  
DB 253 IHTDRQWKVLSKLNKYQDSLFTSPTVARNVAAPLVKIDKVLV-----ADRVSAPE-K 304

QY 297 VLIAGHTNLANLGGALELN-WTLPGQPDNTPPGGELVFERWRRLSDNSQIQLVSLVFQ 355  
DB 305 VTVIVGHDSNIALTLDFKPYQLHDQVETPIGGQLVFRWHDGNAERDLMKIEVYVQ 364

QY 356 TLQQMRDKTFLSLNTPPGVVKLTLAGEERNAGMCSLAGFTQIVNEA 403  
DB 365 SARQLRAEALTLKSPAQRVTLKGCPCV-DANGFCPLDKFDNVNNTA 411

RESULT 3

AGP\_ECOLI ID AGP\_ECOLI STANDARD; PRT; 413 AA.

AC P19926;

DT 01-FEB-1991 (Rel. 17, Created)

DT 01-FEB-1991 (Rel. 17, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Glucose-1-phosphatase precursor (EC 3.1.3.10) (GIPase).

GN AGP OR B1002.

OS Escherichia coli.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI\_TaxID=562;

RN [1]

CC -----

CC EMBL; AE008748; AAL20049.1; --

CC EMBL; U75949; AAC45604.1; --

CC HSP; P07102; IDXL.

CC StyGene; SG10595; agp.

CC InterPro; IPR00560; HisAc\_pheptase.

CC Pfam; PF00328; acid\_phosphat; 1.

CC PROSITE; PS00616; HIS\_ACID\_PHOSPHAT\_1; 1.

CC PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.

CC Hydrolase; Periplasmic; Signal; Complete proteome.

CC SIGNAL 1 22 BY SIMILARITY.

CC FT CHAIN 23 413 GLUCOSE-1-PHOSPHATASE.

CC FT ACT SITE 40 40 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).

CC FT ACT SITE 311 311 PROTON DONOR (BY SIMILARITY).

CC FT CONFLICT 57 57 T -> P (IN REF. 2).

CC FT CONFLICT 60 63 AMPA -> TCPP (IN REF. 2).

CC FT CONFLICT 74 74 K -> T (IN REF. 2).

CC SEQUENCE 413 AA; 45559 MW; E057667C7F8A7244C CRC64;

Query Match 25.5%; Score 548; DB 1; Length 413;  
Best Local Similarity 33.8%; Pred. No. 8.1e-37;  
Matches 138; Conservative 66; Mismatches 180; Indels 24; Gaps 10;

QY 1 QSEPE-LKLESVIVSRHGVRAPIKATQMDVTPDAWTPVKLGELTPRGGLIAYLGH 60  
DB 23 QSEPE-LKLESVIVSRHGVRAPIKATQMDVTPDAWTPVKLGELTPRGGLIAYLGH 82

QY 61 YWRQLVADGILLPKCGCPQSGQVAILADVDERTKGTGEAFAGLAPDCAITVHTQADTSS 120  
DB 83 YWRQLVADGILLPKCGCPQSGQVAILADVDERTKGTGEAFAGLAPDCAITVHTQADTSS 142

QY 121 PDPLFNPLKTVGCOLDNANVTDAILERAGSIADFTGHYQTAFAELERVLNFPQSNLC 180  
DB 143 PDPLFNPLKTVGCOLDNANVTDAILERAGSIADFTGHYQTAFAELERVLNFPQSNLC 202

QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLSMILTEIFLQQAQGMPEPGWRITDS 240  
DB 203 REKQDESCSLTQALPSELKVSADCVSLTGAVSLSMILTEIFLQQAQGMPEPGWRITDS 262

QY 241 HOWNTLILSHNAQFDLLQRTPEVARSRATPLDLIKTALTTPHPKQAYGTLPTSVLFI 300  
DB 263 HOWNTLILSHNAQFDLLQRTPEVARSRATPLDLIKTALTTPHPKQAYGTLPTSVLFI 322

QY 301 AGHDTNLANLGGALELN-WTLPGQPDNTPPGGELVFERWRRLSDNSQIQLVSLVFQTLQOM 360  
DB 323 AGHDTNLANLGGALELN-WTLPGQPDNTPPGGELVFERWRRLSDNSQIQLVSLVFQTLQOM 382

QY 361 RDKTPLSLNTPPGVVKLTLAGEERNAGMCSLAGFTQIVNEAIPACSL 410  
DB 383 RDKTPLSLNTPPGVVKLTLAGEERNAGMCSLAGFTQIVNEAIPACSL 432

RESULT 2

AGP\_SALTY ID AGP\_SALTY STANDARD; PRT; 413 AA.

AC Q33921;

DT 15-JUL-1998 (Rel. 36, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Glucose-1-phosphatase precursor (EC 3.1.3.10) (GIPase).

GN AGP OR STM117.

OS Salmonella typhimurium.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Salmonella.

OX NCBI\_TaxID=602;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=LT2 / SGSC1412 / ATCC 700720;

RX MEDLINE=21534948; PubMed=11677609;

RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P., Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D., Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E., Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M., Waterston R., Wilson R.K.;

RA "Complete genome sequence of Salmonella enterica serovar Typhimurium LT2.";

RT Nature 413:852-856(2001).

RL [2]

RN SEQUENCE OF 1-249 FROM N.A.

RP STRAIN=TN1379;

RC MEDLINE=97405892; PubMed=9260936;

RX Gupta S.D., Wu H.C., Rick P.D.;

RA "A Salmonella typhimurium genetic locus which confers copper tolerance on copper-sensitive mutants of Escherichia coli.";

RL J. Bacteriol. 179:4977-4984(1997).

CC -I- CATALYTIC ACTIVITY: D-glucose 1-phosphate + H(2)O = D-glucose + phosphate.

CC -I- SUBUNIT: Homodimer (By similarity).

CC -I- SUBCELLULAR LOCATION: Periplasmic.

CC -I- SIMILARITY: Belongs to the histidine acid phosphatase family.

CC -----

RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-31.  
RC STRAIN=K12;  
RX MEDLINE=90130318; PubMed=2153660;  
RA Pradel E., Marck C., Boquet P.L.;  
RT "Nucleotide sequence and transcriptional analysis of the Escherichia  
col i agp gene encoding periplasmic acid glucose-1-phosphatase";  
RL J. Bacteriol. 172:802-807(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12";  
RL Science 277:1433-1474(1997).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97061202; PubMed=8905232;  
RA Ohshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
Mori H., Mocomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,  
Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
Yano M., Horiuchi T.;  
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome  
corresponding to the 12.7-28.0 min region on the linkage map";  
RL DNA Res. 3:137-155(1996).  
RN [4]  
RP SEQUENCE OF 23-34.  
RC STRAIN=K12 / EMG2;  
RX MEDLINE=97443975; PubMed=9298646;  
RA Link A.J., Robison K., Church G.M.;  
RT "Comparing the predicted and observed properties of proteins encoded  
in the genome of Escherichia coli K-12";  
RL Electrophoresis 18:1259-1313(1997).  
CC -1- FUNCTION: ABSOLUTELY REQUIRED FOR THE GROWTH OF E. COLI IN A HIGH-  
PHOSPHATE MEDIUM CONTAINING G-1-P AS THE SOLE CARBON SOURCE.  
CC -1- CATALYTIC ACTIVITY: D-glucose 1-phosphate + H(2)O = D-glucose +  
phosphate.  
CC -1- ENZYME REGULATION: INDEPENDENT FROM INORGANIC PHOSPHATE  
AVAILABILITY, AND APPARENTLY SUBMITTED TO CATABOLITE REPRESSION,  
IT IS POSITIVELY CONTROLLED BY CAMP AND THE CAMP RECEPTOR PROTEIN.  
CC -1- SUBUNIT: Homodimer.  
CC -1- SUBCELLULAR LOCATION: Periplasmic.  
CC -1- MISCELLANEOUS: OPTIMALLY ACTIVE AROUND PH 4.  
CC -1- SIMILARITY: Belongs to the histidine acid phosphatase family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M33807; AAA23426.1; -;  
CC EMBL; AE000202; AAC74087.1; -;  
CC EMBL; D90737; BAA35769.1; -;  
CC EMBL; D90738; BAA35779.1; -;  
CC PIR; JVO087; JVO087.  
CC HSSP; P07102; 1DKM.  
CC EcoGene; EG10033; accp.  
CC InterPro; IPR000560; HisAc\_phsphtase.  
CC Pfam; PF00328; acid\_phosphat; 1.  
CC PROSITE; PS00616; HIS\_ACID\_PHOSPHAT\_1; 1.  
CC PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.  
KW Hydrolase; Periplasmic; Signal; Complete proteome.  
FT SIGNAL 1 22  
FT CHAIN 23 413 GLUCOSE-1-PHOSPHATASE.  
FT ACT\_SITE 40 40 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).  
FT -----

FT ACT\_SITE 311 311 PROTON DONOR (BY SIMILARITY).  
SQ SEQUENCE 413 AA; 45683 MW; ADADAD3639D0D6AB CRC64;  
Query Match 25.3%; Score 544; DB 1; Length 413;  
Best Local Similarity 33.1%; Pred.No. 1.7e-36;  
Matches 137; Conservative 71; Mismatches 170; Indels 36; Gaps 12;  
QY 1 QSEPE-LKLESVIVSRHGVRAP-TKATQLMQDVTDPDAPFTWPKVGLGELTPRGELIAYL 58  
DB 23 QTVPEGYQLQVLMMSRHRLRAPLANNQSVLEQSTPNKPEWDVPGGQLITKGGVLEVM 82  
QY 59 GHYWRQRLVADGLLKPCKGCGSQSGQVAILIADVDERTKTRTGEAFAPAGLADPCAITVHTQADT 118  
DB 83 GHYVREWLAEQGMVKGSGCEPPYTVYAYANSLQRTVATRAQFFITGAPFGCDIPVHHQEKM 142  
QY 119 SSPDPLFNPLKTVGCQLDNANVTD---AIIERAGGSIADFTGHYQTAFRELRLVNFPOS 175  
DB 143 GTMDPTFNPVITD---DSAAFSEQAVAMEKLSKL-QLTDSVQL---LEKIVNVKDS 193  
QY 176 NLCLKREKQDESCSLTQALPSELKVSADCVSLTQAVSLASMLTEIFILQQAQGNP--EPG 233  
DB 194 PAC---KEKQCCSLVDGKNTFSAYQQEFGVSGFLKVGNSLVDAFTLQYYEGFPMQVA 249  
QY 234 WGRITDSHOWNTLLSLHNAQFDLLQRTPEVARSRATPLLDLIKLTALTTPHPPQKQAYVTL 293  
DB 250 WGEIKSDQWKVLSKXGYQDSLFTSFVARNVAKLVSYIDKAL-----VTD 298  
QY 294 PTS---VLFIAGHDTNLANLGALLEN-WTLPGGPDNTPPGGLVFFERWRRLSDNSQWIQ 349  
DB 299 RTSAPKITVLVGHDSNLSIALTLTALDFKPYQLHDQNETPIGGKIVFQVHWDHDSKANRDLMK 358  
QY 350 VSLVPQTLQOMRDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGTQIVNEA 403  
DB 359 IEYVQSAEQURNADALTQAPORVILELGGC-PIDADGFCPMKDFSDVLENEA 411  
RESULT 4  
ID AGP PROBE STANDARD; PRT; 417 AA.  
AC Q52309.  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Glucose-1-phosphatase precursor (EC 3.1.3.10) (Glpase).  
GN AGP.  
OS Providencia rettgeri.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Providencia.  
OX NCBI\_TaxID=597;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PV7;  
RA Riccio M.L., Chiesurin A., Lombardi G., Satta G.;  
RL Submitted (SEP-1992) to the EMBL/GenBank/DBJ databases.  
CC -1- CATALYTIC ACTIVITY: D-glucose 1-phosphate + H(2)O = D-glucose +  
phosphate.  
CC -1- SUBUNIT: Homodimer (By similarity).  
CC -1- SUBCELLULAR LOCATION: Periplasmic (By similarity).  
CC -1- SIMILARITY: Belongs to the histidine acid phosphatase family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X68201; CAA48288.1; -;  
CC PIR; S25627; S25627.  
CC HSSP; P07102; 1DKL.  
CC InterPro; IPR000560; HisAc\_phsphtase.  
CC Pfam; PF00328; acid\_phosphat; 1.  
CC -----



FT CARBOHYD 267 267 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 29 29 Q -> R (IN REF. 1 AND 2).  
SQ SEQUENCE 423 AA; 48316 MW; 1A2C8FBA0ED3724B CRC64;  
Query Match 6.1%; Score 132; DB 1; Length 423;  
Best Local Similarity 23.3%; Pred. No. 0.0037;  
Matches 97; Conservative 43; Mismatches 173; Indels 104; Gaps 20;  
Qy 8 LESVIVSRHGVRAPTKATQMDVTPDAPWTPVKLGELTPRGELIAYLGHYWRQLV 67  
Db 33 LRFVTLRYHGRSPVKT---YPKDPYQBEWPGQFGLTKGMLQHWELGQALRQY- 87  
Qy 68 ADGLLPKCGCQSQGVAIIADVD-ERTRKTEGFAAGLAPDCAITVHTQADTSSPDPL- 124  
Db 88 -HGFL---NTSYHRQEVVVRSTDFDRLMSAEANLAGLFP-----PQMGQR 129  
Qy 125 FNP-LKTVGCOLDNANVTDAILERAGGSIAFTGHYQTAFRELVNFPQSNLCKREK 183  
Db 130 FNPENISWQIPVHTVPIE-----DRLLKFPGLGPCPRYEQL 165  
Qy 184 ODESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFL-----LQQAQMGPEP 232  
Db 166 QNE-----TRQTPETQNESRNNAQFLDMVANETGLDITLVNVTYTLFCEQTHGLRIP 221  
Qy 233 GWGRITDSHOWNTLLSLHNAQFDLL---QRTPEVARSRATPLLDLTKALTTPHPQKQAY 289  
Db 222 PW---ASPQTMORLSRLKDFSRFLFGIYQAEKARLQGGVLLAQIKNLTLWATTSQ-- 276  
Qy 290 GVTLPTSVLFTAGHDTNLANLGGALELNLWLPQGPDPNTPGGELVPERWRRLSDNSQWIQ 349  
Db 277 ---LP-KLLVYSADHTTLVALQMALDVI-----NGEQAPYASCHIFFELYQEDSGN---FS 324  
Qy 350 VSLVFQTLQOQRDKTPLSLNTPPGEVKLTLAGCEERNAQMGCSLAGTQIVNEARIP 406  
Db 325 VEMTF---RNESDKAPWPLSLP-----GCPRH-----CPLQDFLRL-TEPVVP 363  
RESULT 6  
FPAL RAT STANDARD; PRT; 423 AA.  
AC P20611;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE Lysosomal acid phosphatase precursor (EC 3.1.1.3.2) (LAP).  
GN ACP2.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89350910; PubMed=2764916;  
RA Hienno M., Fujita H., Noguchi Y., Kono A., Kato K.;  
RT "Isolation and sequencing of a cDNA clone encoding acid phosphatase  
in rat liver lysosomes";  
RL Biochem. Biophys. Res. Commun. 162:1044-1053(1989).  
CC -!- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an  
alcohol + phosphate.  
CC -!- SUBCELLULAR LOCATION: Lysosomal.  
CC -!- SIMILARITY: Belongs to the histidine acid phosphatase family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.ebi.ac.uk/ebis-sib.ch/>)  
CC or send an email to [license@ebi-sib.ch](mailto:license@ebi-sib.ch).  
CC  
CC EMBL; M27893; AAA40744.1; .

DR A33395; A33395.  
DR HSP; E20446; IRPA.  
DR InterPro; IPR000560; HisAc\_phosphatase.  
DR Pfam; PF0328; acid\_phosphat; 1.  
DR PROSITE; PS00616; HIS\_ACID\_PHOSPHAT\_1; 1.  
DR PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.  
KW Hydrolase; Signal; Glycoprotein; Lysosome.  
FT SIGNAL 1 30  
FT CHAIN 31 423 LYSOSOMAL ACID PHOSPHATASE.  
FT DISULFID 159 370 BY SIMILARITY.  
FT DISULFID 212 310 BY SIMILARITY.  
FT DISULFID 345 349 BY SIMILARITY.  
FT ACT SITE 42 42 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).  
FT ACT SITE 84 84 BY SIMILARITY.  
FT ACT\_SITE 286 286 PROTON DONOR (BY SIMILARITY).  
FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 167 167 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 177 177 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 197 197 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 267 267 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 423 AA; 48319 MW; 6F4C4819CADD4496 CRC64;  
Query Match 5.3%; Score 114; DB 1; Length 423;  
Best Local Similarity 22.7%; Pred. No. 0.11;  
Matches 95; Conservative 42; Mismatches 175; Indels 106; Gaps 20;  
Qy 8 LESVIVSRHGVRAPTKATQMDVTPDAPWTPVKLGELTPRGELIAYLGHYWRQLV 67  
Db 33 LRFVTLRYHGRSPVKA---YPKDPYQBEKWPQGFGLTKGMLQHWELGQALRQY- 87  
Qy 68 ADGLLPKCGCQSQGVAIIADVD-ERTRKTEGFAAGLAPDCAITVHTQADTSSPDPLFN 126  
Db 88 -HGFL---NASYHRQEVVVRSTDFDRLMSAEANLAGLFP-----TEVQHFNISWQ 137  
Qy 127 PLKTVGCOLDNANVTDAILERAGGSIAFTGHYQTAFRELVNFPQSNLCKREKQDE 186  
Db 138 PT-----FVHTVPIE-----DRLLKFPGLGPCPRYEQLQNE 168  
Qy 187 SCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFL-----LQQAQMGPEPQWG 235  
Db 169 ---TRQTPETQNESRNNAQFLDMVANETGLMNLTLNVTYTLFCEQTHGLLPPW- 223  
Qy 236 RITDSHOWNTLLSLHNAQF-----DLLQRTPEVARSRATPLLDLTKALTTPHPQKQA 288  
Db 224 ---ASPQTMORLSRLKDFSRFLFGIYQAEKARLQGGVLLAQIKNLTLWATTSQF 277  
Qy 289 YGVTLPTSVLFTAGHDTNLANLGGALELNLWLPQGPDPNTPGGELVPERWRRLSDNSOWI 348  
Db 278 ---PKLLVYSADHTTLVALQMALNVI-----NGKQAPYASCHIFFELYQ--EDNGNF- 323  
Qy 349 QVSLVFQTLQOQRDKTPLSLNTPPGEVKLTLAGCEERNAQMGCSLAGTQIVNEARIP 406  
Db 324 SVEMTFRNDSK---KAPW-----PLTLPGCPHR-----CPLQDFLRL-TEPVIP 363  
RESULT 7  
PAX8\_MOUSE STANDARD; PRT; 457 AA.  
ID PAX8\_MOUSE  
AC Q00288; P97342;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Paired box protein Pax-8.  
GN PAX8 OR PAX-8.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]





EMBL; L02420; AAA16897.1; -.  
PIR; JN0890; JN0890.  
PDB; 1QFX; 19-APR-00.  
InterPro; IPR000560; HisAc\_phspstee.  
Pfam; PF00328; acid\_phosphat\_1.  
PROSITE; PS00616; HIS\_ACID\_PHOSPHAT\_1.  
PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2.  
Hydrolase; Glycoprotein; Signal; 3D-structure.  
SIGNAL 1 49  
FT CHAIN 20 479  
FT ACT\_SITE 82 82  
FT ACT\_SITE 337 337  
FT ACT\_SITE 337 337  
FT DISULFID 71 387  
FT DISULFID 128 472  
FT DISULFID 216 441  
FT DISULFID 225 298  
FT DISULFID 413 421  
FT CARBOHYD 191 191  
FT CARBOHYD 315 315  
FT CARBOHYD 458 458  
FT HELIX 41 43  
FT HELIX 45 47  
FT TURN 50 51  
FT TURN 69 70  
FT STRAND 71 81  
FT STRAND 85 85  
FT HELIX 89 103  
FT HELIX 112 119  
FT HELIX 127 129  
FT TURN 130 131  
FT STRAND 133 133  
FT TURN 138 139  
FT HELIX 141 156  
FT HELIX 157 159  
FT STRAND 166 171  
FT HELIX 173 173  
FT HELIX 174 188  
FT HELIX 189 191  
FT STRAND 192 195  
FT HELIX 205 207  
FT TURN 208 209  
FT TURN 224 227  
FT TURN 233 233  
FT HELIX 234 245  
FT TURN 247 248  
FT HELIX 253 266  
FT TURN 267 269  
FT TURN 274 275  
FT HELIX 276 278  
FT HELIX 281 298  
FT TURN 301 302  
FT TURN 304 305  
FT HELIX 306 323  
FT HELIX 325 328  
FT STRAND 331 336  
FT HELIX 338 348  
FT TURN 349 349  
FT TURN 364 365  
FT HELIX 370 372  
FT TURN 373 373  
FT TURN 376 377  
FT STRAND 379 387  
FT STRAND 390 390  
FT TURN 391 392  
FT STRAND 393 393  
FT STRAND 396 403  
FT TURN 404 405  
FT STRAND 406 408  
FT STRAND 410 410  
FT TURN 411 412  
FT STRAND 413 413  
FT HELIX 417 419

Query Match 5.1%; Score 109.5; DB 1; Length 479;  
Best Local Similarity 19.2%; Pred. No. 0.29;  
Matches 91; Conservative 48; Mismatches 145; Indels 189; Gaps 23;  
7 KLESVIVSRHGVRAPTKATQLMQDVTDPDAMPVFKL-----GELT----- 48  
72 EVDQVIMVKRHRGERVPSAG--KDIEEALAKYSINTTEYKGDLPFLNDWTVVPEY 129  
49 -----PRGELIAY-----LGHYRQRLVADGLLPKCGCPQSGVAILADVDE 91  
130 YNAETTSQPYAGLLDAYNHGNDYKARYGHLWNGETVVFF-----SSGYGVI--E 178  
92 RTRKTGEA-FAAGLAPDCAITVHTQADTSSPDPLFNPLKTVGVQOLDNANVTDAILERAGG 150  
179 TARKEGEGFGYNYSTNAALNIISEEVMGADSL-----IPTCDTNDQTCNLT--- 229  
151 STADFTGHYQAFRELERVLPQSNLCIKR-EKQDESCSILTQALPSELKVSADCVSLTG 209  
230 -----YQ-----LPOFKVAARLNSQNPQMLT-----ASDVNVL-- 259  
210 AVSLASMLTEIFLLQQAQMPGPGWGRITDISHQNTL-----LSLH----- 250  
260 -MVMAF-----ELNARFFSNWNAFTQDENVSFGYVEDLNYYCAGPGDKNMAAVG 310  
251 -----NAQFDLLQRTPEVARSRATPLLDLIKLTALTPHPQKQAYGVTLPTSVLFIAGHDTN 306  
311 AVYANASUTLLNQGEKA-----GSLFFNFADHTN 340  
307 LANLGGALEL---NWTLP-----GOP-----DNTFPGGELVFERW-----RRLSDNSQW 349  
341 ITPILAALGVLPINEDLPDRVAFGNPVYSIGNIVPMGGHLTIERLSQCATLSDEGTYVR 400  
350 VSLVFQTLQWRDKTPLSLNTPPGCVKLTLAGCERNNAQNCISLAGTQIVNE 402  
401 LVL-----NEAVLFNDCTSGPGYS-CPLANYTTSILNK 432

RESULT 9  
RY444 DROME  
ID\_RY444 DROME STANDARD; PRT; 5127 AA.  
AC Q24498; Q24321; Q24499; Q24500; Q24501; Q8MKS3; Q8MKS4; Q8MKS5;  
AC Q9V4Y7;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Ryanodine receptor 44F.  
GN RYA-R44F OR DRY OR RYR OR CG10844.  
OS Drosophila melanogaster (Fruit fly).  
CC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
CC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS A; B; C AND D), FUNCTION, TISSUE  
RP SPECIFICITY, AND DEVELOPMENTAL STAGE.  
RX MEDLINE=94102409; PubMed=8276118;  
RA Takeshima H., Nishi M., Iwabe N., Miyata T., Hosoya T., Masai I.,  
RA Hotta Y.;  
RT "Isolation and characterization of a gene for a ryanodine  
RT receptor/calcium release channel in Drosophila melanogaster."  
RL FEBS Lett. 337:81-87(1994).  
RN [2]

RP SEQUENCE FROM N.A.  
RC STRAIN=Berkeley;  
RX MEDLINE=2019606; PubMed=10731132;  
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Gallie R.F.,  
RA George R.A., Lewis S.B., Richards S.J., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
RA Ballew R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fiolier C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D.A., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinart K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Wang X.,  
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasarman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RA "The genome sequence of Drosophila melanogaster.";  
RT Science 287:2185-2195(2000).  
[3]  
RP REVISIONS, AND ALTERNATIVE SPLICING.  
RX MEDLINE=2242606; PubMed=12537572;  
RA Miera S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradecky P., Huang Y., Kaninker J.S., Millburn G.H., Prochnik S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Bettencourt B.R., Celisner S.E., de Grey A.D.N.J., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shy S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
RT systematic review".  
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
[4]  
RP SEQUENCE OF 4619-5127 FROM N.A., DEVELOPMENTAL STAGE, AND TISSUE  
SPECIFICITY.  
RX MEDLINE=93202018; PubMed=1338312;  
RA Haan G., Roobash M.;  
RT "Drosophila homologs of two mammalian intracellular Ca(2+)-release  
RT channels: identification and expression patterns of the inositol  
RT 1,4,5-trisphosphate and the ryanodine receptor genes.";  
RL Development 116:967-975(1992).  
[5]  
RP FUNCTION, AND TISSUE SPECIFICITY.  
RX MEDLINE=20283930; PubMed=10811919;  
RA Sullivan K.M.C., Scott K., Zuker C.S., Rubin G.M.;  
RT "The ryanodine receptor is essential for larval development in  
RT Drosophila melanogaster.";  
RA Proc. Natl. Acad. Sci. U.S.A. 97:5942-5947(2000).  
CC -!- FUNCTION: Communication between transverse-tubules and  
CC sarcolemmal reticulum. Contraction of muscle is triggered by

CC release of calcium ions from SR following depolarization of T-  
CC tubules (By similarity).  
CC -!- FUNCTION: Intracellular calcium channel that is required for  
CC proper muscle function during embryonic development and may be  
CC essential for excitation-contraction coupling in larval body wall  
CC muscles.  
CC -!- SUBUNIT: Homotetramer (Probable).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=4;  
CC Name=A;  
CC IsoId=Q24498-2; Sequence=Displayed;  
CC Name=B;  
CC IsoId=Q24498-3; Sequence=VSP\_050196;  
CC Name=C;  
CC IsoId=Q24498-4; Sequence=VSP\_050195;  
CC Name=D;  
CC IsoId=Q24498-5; Sequence=VSP\_050195, VSP\_050196;  
CC TISSUE SPECIFICITY: During embryonic stages 9-10, expression is  
CC seen in mesoderm of all segments in progenitors of the cephalic  
CC and somatic muscles. Adults exhibit high expression in tubular  
CC 'jump' muscles of thorax and leg, and lower expression in the  
CC brain, ventral ganglion, head muscles and proboscis muscles.  
CC -!- DEVELOPMENTAL STAGE: Abundant in 6-12 hour embryos, reduced  
CC expression in second and third instar larval stages and adults.  
CC -!- MISCELLANEOUS: The calcium release channel is modulated by calcium  
CC ions, magnesium ions, ATP and calmodulin (By similarity).  
CC -!- MISCELLANEOUS: The calcium release channel activity resides in the  
CC C-terminal region while the remaining part of the protein  
CC constitutes the 'foot' structure spanning the junctional gap  
CC between the SR and the T-tubule. It is possible that the foot  
CC structure interacts with the cytoplasmic region of the  
CC dihydropyridine receptor (By similarity).  
CC -!- MISCELLANEOUS: Ryanodine is an alkaloid that binds to the Ca-  
CC release channel in junctional SR and modulates its activity (By  
CC similarity).  
CC -!- SIMILARITY: Belongs to the ryanodine receptor family.  
CC -!- SIMILARITY: Contains 5 MIR domains.  
CC -!- SIMILARITY: Contains 3 SPRY domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; D17389; BAA04212.1; -  
CC EMBL; D17389; BAA41469.1; -  
CC EMBL; D17389; BAA41470.1; -  
CC EMBL; D17389; BAA41471.1; -  
CC EMBL; AE003835; AAF59036.2; -  
CC EMBL; AE003835; AAM71082.1; -  
CC EMBL; AE003835; AAM71083.1; -  
CC EMBL; AE003835; AAM71084.1; -  
CC EMBL; Z18536; CAA79221.1; -  
CC FlyBase; FBgn0011286; Rya-r44f.  
CC GO; GO:0016021; C:integral to membrane; NAS.  
CC GO; GO:0005219; P:ryanodine-sensitive calcium-release channel. . .; TAS.  
CC GO; GO:0006816; P:calcium ion transport; TAS.  
CC GO; GO:0006936; P:muscle contraction; IMP.  
CC InterPro; IPR000699; Ca-rel\_channel.  
CC InterPro; IPR001682; Ca/Na\_Pore.  
CC InterPro; IPR002048; EF-hand.  
CC InterPro; IPR005821; Ion\_trans.  
CC InterPro; IPR003608; MIR.  
CC InterPro; IPR001215; Ryanodn\_receptor.  
CC InterPro; IPR003032; RYR.  
CC InterPro; IPR003877; SPRY\_receptor.  
CC Pfam; PF00520; Ion\_trans; 1.  
CC Pfam; PF02815; MIR; 4.  
CC Pfam; PF01365; RYDR\_ITPR; 2.

DR Pfam; PF002026; RYR; 4.  
DR Pfam; PF00622; SPRY; 3.  
DR PRINTS; PR00795; RYANODINER.  
DR SMART; SM00472; MIR; 4.  
DR SMART; SM00449; SPRY; 3.  
DR PROSITE; PS00919; MIR; 5.  
KW Repeat; Transmembrane; Receptor; Ionic channel; Calcium channel;  
KW Developmental protein; Glycoprotein; Phosphorylation;  
KW Alternative splicing.  
FT DOMAIN 1 3232  
FT TRANSNEM 3233 3253  
FT TRANSNEM 3201 3321  
FT TRANSNEM 3376 3396  
FT TRANSNEM 3963 3983  
FT TRANSNEM 4452 4472  
FT TRANSNEM 4651 4671  
FT TRANSNEM 4683 4703  
FT TRANSNEM 4742 4762  
FT TRANSNEM 4885 4905  
FT TRANSNEM 4924 4944  
FT TRANSNEM 4963 4983  
FT TRANSNEM 5004 5024  
FT DOMAIN 94 148  
FT DOMAIN 155 200  
FT DOMAIN 210 264  
FT DOMAIN 270 328  
FT DOMAIN 336 394  
FT DOMAIN 663 802  
FT DOMAIN 1091 1216  
FT DOMAIN 1529 1672  
FT DOMAIN 4686 4711  
Query Match 5.18; Score 109; DB 1; Length 5127;  
Best Local Similarity 20.98; Pred. No. 7.8;  
Matches 89; Conservative 45; Mismatches 133; Indels 158; Gaps 18;  
QY 42 VKLGELTPRGEL---IAYIGHYWRQLVA-----DGLPKCGCPSGGVAIIADVDT 93  
Db 3934 LNLGAILRGNNIDIQMGLNHLKXKDVGFSTIAGLWNSCS-----VLDLDAFE 3984  
QY 94 RKTGEAPAGLAPCALVHTQADTSPPDLF-----NPLKTGCQCLDNA 138  
Db 3985 RNT-KASGLGVSGAAGERKMDHDETCALRFIQTCEGHNLWQNYLRTQAGNTTV 4043  
QY 139 NV---TDAILERAGGSIAFTGHYQTAFLERLERNVFNPSQNLCLKREKODESCSLTQALP 195  
Db 4044 NVVICTVYLLRLQESIMDFVWHYSS-----LLEIF-----LLQOAGWPEPGWGIITDSHOMNTL 4069  
QY 196 SELKVSADCVSLTGAVALSM-----LLEIF-----LLQOAGWPEPGWGIITDSHOMNTL 246  
Db 4070 KEIIDPAGKANFFKAIIVASQVFNLTETVIGQPTLNQQA-----LAHSLRLDAV 4119  
QY 247 -----LSLENAQFDLLQRTPEVARSRAATPLLDLTKTALTPHPPOKQAYGVTL 293  
Db 4120 GGFLFLFSHMQDKLSKSSQVDLLKE-----LLNLQKDMIT----- 4155  
QY 294 PTVSLFAGHTDTNLANLGGALELNTWLPQGPDNTPPGGELV---FEWRRLSD----- 343  
Db 4156 ---MLSMLEGVNVGTIGKQVDDTLVESASNV-----ELILKYDFMLKLADIESPSF 4207  
QY 344 -----NSQMTQVSLVFTQLQMRDKTFLSLNTPFGVGLKTLACGERNAQGMCSLAGFT 397  
Db 4208 HEVDMKNEGIV-----TPKDFREKMEQSKNYTPEEMDFILACC-ERNHEGKIDYRAFV 4259  
QY 398 QIVNE 402  
Db 4260 EHFHE 4264  
RESULT 10  
PHYB\_ASPNG STANDARD; PRT; 479 AA.  
ID PHYB\_ASPNG  
AC P34754;

DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE 3-phytase B precursor (EC 3.1.3.8) (Myo-inositol-hexaphosphate  
DE 3-phosphohydrolase B) (3 phytase B) (Myo-inositol hexakisphosphate  
DE Phosphohydrolase B).  
GN PHYB.  
OS Aspergillus niger.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eutotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
OX NCBI\_TaxID=5061;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 20-101; 133-141 AND 376-399.  
RX MEDLINE=93371452; PubMed=7916610;  
RA Ehrlich K.C., Montalbano B.G., Mullaney E.J., Dischinger H.C. Jr.,  
RA Ullah A.H.J.;  
RT "Identification and cloning of a second phytase gene (phyB) from  
RT Aspergillus niger (ficusum).";  
RL Biochem. Biophys. Res. Commun. 195:53-57(1993).  
CC -!- FUNCTION: Catalyzes the hydrolysis of inorganic orthophosphate  
CC from phytate.  
CC -!- CATALYTIC ACTIVITY: Myo-inositol hexakisphosphate + H(2)O = 1D-  
CC myo-inositol 1,2,4,5,6-pentakisphosphate + phosphate.  
CC -!- SIMILARITY: Belongs to the histidine acid phosphatase family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC EMBL; L20567; AAA02934.1; -.  
DR HSP; P34755; IQPX.  
DR InterPro; IPR000560; HisAc\_phosphatase.  
DR Pfam; PF00328; acid\_phosphat; 1.  
DR PROSITE; PS00616; HIS\_ACID\_PHOSPHAT\_1; 1.  
DR PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.  
KW Hydrolase; Glycoprotein; Signal.  
FT SIGNAL 1 19  
FT CHAIN 20 479  
FT ACT\_SITE 81 81  
FT ACT\_SITE 82 82  
FT ACT\_SITE 337 337  
FT CARBOHYD 106 106  
FT CARBOHYD 191 191  
FT CARBOHYD 227 227  
FT CARBOHYD 250 250  
FT CARBOHYD 315 315  
FT CARBOHYD 425 425  
FT CARBOHYD 442 442  
FT CARBOHYD 458 458  
FT SEQUENCE 479 AA; 52611 MW; 395D4DA2B50PDFC4 CRC64;  
Query Match 5.0%; Score 108.5; DB 1; Length 479;  
Best Local Similarity 20.2%; Pred. No. 0.35;  
Matches 95; Conservative 44; Mismatches 147; Indels 185; Gaps 25;  
QY 7 KLESVIVSRHGRVAPT-----KATQMDVT-PDAW----PTWPVKL 44  
Db 72 EVDQVIMVKRGERYPSAGKSIEELAKVYSINTYKGLAFNDWTYVVPNECYN 131  
QY 45 GELT--PRGELIAY-----LGHYWRQLVADGLLPKCGCPQSQVAILADVDT 93  
Db 132 AETTSGPYAGLLDAYNHGNDYKARYGHLWNGETVVVFF-----SSGYGRVI-----ETA 180  
QY 94 RKTGEA-FAAGLAPDCAITVHTQADTSPPDLNPLNFKTGVQCLDNVNTDAILERAGGS 152  
Db 181 RKFGSGFGYNYSTNAALNIISEVWGADSL-----TPTCDTNDQTTCDNLT----- 229  
QY 153 ADFTGHYQTAFLERLERNVFNPSQNLCLKREKODESCSLTQALPSELKVSADCVSLTGA 211

```
Db 230 -----YQ-----LPQKVAARLNSQNGHNTL-----ASDVNLT--I 260
QY 212 SLASMLTEIFLLQQAQMPFGNGRITDTSQHWNTL-----LSLH----- 250
Db 261 VWASF-----ELNARPPFNWNAFTQDEWVSFGVVEDLNYYCAGPGDKNMAAVGAV 312
QY 251 --NAQFDLQRTPEVARSRAFLDLIKALTALTPHPQKQAVGVTLPVSLFIAGHDNTLA 308
Db 313 YANASLTLLNGPK-----EAGPL-----FFNFAHDNTIT 342
QY 309 NLGGALEL---NWTLP-----GQP-----DNTPPGGELVFERM-----RRLSNDSQNTQVS 351
Db 343 PILAALGVLIENEDLFLDRVAFNGPYSGNIVPMGGHLLTIERLSQATALSDDKGYVRLV 402
QY 352 LVFQTLQMRDKTPLSLNTPGVEVKLILAGCEERNAQMGSLAGFTQIVNE 402
Db 403 L-----NEAVLPNDCTSGFGYS-CPLANYTSILNK 432

RESULT 11
PAX8_CANFA
ID PAX8 CANFA STANDARD; PRT; 459 AA.
AC P47240; P47241;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Paired box protein Pax-8.
GN PAX8 OR PAX-8.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 8A AND 8G).
RC TISSUE=Thyroid;
RX MEDLINE=96254068; PubMed=8652674;
RA van Renterghem P.H.G., Vassart G., Christophe D.;
RT "Pax 8 expression in primary cultured dog thymocyte is increased by
cyclic AMP.";
RL Biochim. Biophys. Acta 1307:97-103(1996).
CC -!- FUNCTION: THOUGHT TO ENCODE A TRANSCRIPTION FACTOR. IT MAY HAVE A
MAMMALIAN DEVELOPMENT.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=8A;
CC IsoId=P47240-1; Sequence=Displayed;
CC Name=8G;
CC IsoId=P47240-2; Sequence=VSP 002371;
CC -!- SIMILARITY: Contains 1 paired box domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X83591; CAA58571.1; -.
CC EMBL; X83592; CAA58572.1; -.
CC PIR; S70361; S52250.
CC PIR; S70362; S52251.
CC HSSP; P26367; 6PAX.
CC TRANSFAC; T02930; -.
CC InterPro; IPR001523; Paired_box.
CC Pfam; PF00292; PAX; 1.
CC PRINTS; PR00027; PAIREDBOX.
CC SMART; SM00351; PAX; 1.
CC PROSITE; PS00034; PAIRED_BOX; 1.
```

```
KW Paired box; Developmental protein; Nuclear protein; DNA-binding;
KW Transcription regulation; Differentiation; Alternative splicing.
FT DOMAIN 9 133
FT VARSPLIC 302 459
FT -----
FT PAIRED BOX.
FT DPHSPFAIKQETPEVSSSSSTPSSLSAFLDLQVGGAGV
FT PAGASVPPFNAPPAASVYQFTGQALLSGREMTPLPGY
FT PPHIPTSGGVSASAIAGVAGSEYGNAYGHTVSYSGE
FT AMWFNSLLSSPYVYSTSRPSAPPTATADFHL -> AP
FT PEMICKSNAQSGQVPSRPSMPMLPCTGSSRRPSS
FT QSERWDEHCDTHTPPVDRAANPLLPDQAWQVPHIT
FT VPHQGVRRHPPRPRLTTCSCYDGSN (in isoform
FT 8G).
FT FTId=VSP 002371.
SQ SEQUENCE 459 AA; 48839 MW; 4CC00164ABE56935 CRC64;

Query Match 5.0%; Score 107.5; DB 1; Length 459;
Best Local Similarity 18.7%; Pred.No.0.4; 130; Indels 171; Gaps 16;
Matches 81; Conservative 51; Mismatch 130; Indels 171; Gaps 16;

QY 4 PELKESVIVSRHGVRAPTKATL-----MDVTPDAWPTWPKLGL-----TP 49
Db 27 PEVVRQRIVDLAHQGVPCDISRQLRVSHGCVSKILGRYYETGSRPGVIGGSKPKVATP 86
QY 50 RGELIAYLGHYWRQ-----RLVADGL-----LP 73
Db 87 K---VVEKIGDYKRONPTWFAWEIRDLAEGVCDNTVPSVSSINRIIRTKVQPFNL 143
QY 74 KGCQ-----POS-----GQVATIAVDTRTRKTGEAF 100
Db 144 MDSCVATKSLSPGHTLIPSSAVTPPESQSDSLGTSYINGLIGTAQPSDSKRKWDSD 203
QY 101 AGLAPDCAITVHTQADRSSDPLFNPLKTGVQCLDNANVTDAILERAGGSTADFTGHVQ 160
Db 204 ----QDSCRLSIDSSSSGPR---KHLRTDAFSQHLEPLECPFERQ-----HYP 247
QY 161 TAPRELERYLNFQSNLCLKREKQDESCLTQALPSLKVSDCVSL-----TGAV 211
Db 248 EAY-----ASPSHKGEQGLYPLPLNSAALDDGKA 278
QY 212 SLASMLTEIFLLQQAQMPFGNGRITDTSQHWNTLHNAQFDLLQRTPEVARSRATP- 270
Db 279 TLTPSNTP-----GNLSTHOTYPPVADPHSPFAIKQETPEVSSSSSTPS 324
QY 271 -----LLDLIKALTALTPHPQKQAVGVTLPVSLF-----IAGHDNTLANLGALELNT 319
Db 325 SLSSSAFLDLQVGGAGVPA-----GASVPPNAPPAASVYQFTGQALLSGREMT 379
QY 320 LFGQPDNTPPGGE 332
Db 380 LFGYPPHIPTSGQ 392

RESULT 12
PAX8_HUMAN
ID PAX8_HUMAN STANDARD; PRT; 450 AA.
AC Q06710; Q09155; Q16337; Q16338; Q16339;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Paired box protein Pax-8.
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND VARIANT LEU-329.
RC TISSUE=Kidney;
RX MEDLINE=93170167; PubMed=1337742;
RA Poleev A., Fickenscher H., Mundlos S., Winterpacht A., Zabel B.,
RA Fidler A., Gruss P., Plachov D.;
RT "PAX8, a human paired box gene: isolation and expression in
developing thyroid, kidney and Wilms' tumors.";
RL Development 116:611-623(1992).
```



Query Match 4.9%; Score 105; DB 1; Length 450;  
Best Local Similarity 18.6%; Pred. No. 0.62;  
Matches 80; Conservative 51; Mismatches 125; Indels 174; Gaps 17;  
QY 4 PELKESVIVSRHGVRAPTKATQL-----MQDVTDPAMTPWPKLGL-----TP 49  
DB 27 PEWQRIVDLAHQGVPCDISRLRVSHGCVSKILGRYETGSRPGVIGSGKPKVATP 86  
QY 50 RGELIAYLGHWRQ-----RLVADGL-----LP 73  
DB 87 K--VVEKIDYKRONPTWFAWEIRLLAEGVCDNTVPVSSINRIIRTKVQPFNLP 143  
QY 74 KCGC-----POSGQVAIIADVDERTRKTGEFA-----AGLA----- 105  
DB 144 MDCSVATKSLSPCHTLTPSS---AVTPSSPOSDSLGTYSLINGLLGIAQPGSKDKKXMD 200  
QY 106 ---PDAITVHQADTSSPDPLFNPLKTVGCOLDNANVTDAILERAGGSIADFTGHYQTA 162  
DB 201 SDQDSCLSIDSQSSSGPR---KHLRTDAFSGHLEPLECPFERQ-----HYPEA 248  
QY 163 PRELERNVLPQSNCLKREKQDECSLTQALPSELKVSADCVSL-----TGAVSILA 214  
DB 249 Y-----APSHTKGEQGLYPLPLNLTSLDDCKATIT 279  
QY 215 SMLTEIFLLQQAQGMPEPCWGRITDSHOWNTLLSHNAQFDLLQRTPEVARSRATP----- 270  
DB 280 PSNTPL-----GRNLSHTQYVPVADPHSPFAIKQETPEVSSSSSTPSSLS 325  
QY 271 ---LRLIKTALTPHPQQAQGVTLPTSFLF-----IAGHDTNLANLGGALELAWTLPG 322  
DB 326 SSAFLDL-----QQVSGVPPFNAPPHAASVYQGTGALLSGREWVGTPLPG 373  
QY 323 QPDNTPPGGE 332  
DB 374 YPHIPTSGQ 383  
RESULT 13  
PPAL MOUSE  
ID PPAL MOUSE STANDARD; PRT; 423 AA.  
AC P24638; Q8Q0T5;  
DT 01-MAR-1992 (Rel. 21, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Lysosomal acid phosphatase precursor (EC 3.1.3.2) (LAP).  
GN ACP2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
SEQUENCE FROM N.A.  
RP TISSUE=Liver;  
RC MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.P., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Guraratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahney J., Heiton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
"Generation and initial analysis of more than 15,000 full-length human

and mouse cDNA sequences.";  
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
[2]  
SEQUENCE OF 3-423 FROM N.A.  
MEDLINE=91282986; PubMed=2059337;  
RA Geier C., von Figura K., Pohlmann R.;  
RT "Molecular cloning of the mouse lysosomal acid phosphatase.";  
Biol. Chem. Hoppe-Seyler 372:301-304(1991).  
CC -1- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an  
alcohol + phosphate.  
CC -1- SUBCELLULAR LOCATION: Lysosomal.  
CC -1- SIMILARITY: Belongs to the histidine acid phosphatase family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/annouce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
EMBL; BC023343; AAH23343.1; -.  
EMBL; X57199; CAA40485.1; -.  
PIR: S14742; S14742.  
HSSP; P15309; ZHPA.  
MGD; MGI:87882; Acp2.  
InterPro; IPR00560; HisAc phsphtse.  
DR Pfam; PF00328; acid phosphat; 1.  
DR PROSITE; PS00616; HIS ACID PHOSPHAT 1; 1.  
DR PROSITE; PS00778; HIS ACID PHOSPHAT 2; 1.  
KW Hydrolase; Signal; Glycoprotein; Lysosome.  
FT SIGNAL 1 30  
FT CHAIN 31 423 LYSOSOMAL ACID PHOSPHATASE.  
FT DISULFID 159 370 BY SIMILARITY.  
FT DISULFID 212 310 BY SIMILARITY.  
FT DISULFID 345 349 BY SIMILARITY.  
FT ACT SITE 42 42 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).  
FT ACT SITE 84 84 BY SIMILARITY.  
FT ACT SITE 286 286 PROTON DONOR (BY SIMILARITY).  
FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 167 167 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 177 177 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 197 197 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 267 267 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 374 374 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 131 131 N -> S (IN REF. 2).  
SQ SEQUENCE 423 AA; 48508 MW; AFD7A5C90A4F2AF4 CRC64;  
Query Match 4.8%; Score 104; DB 1; Length 423;  
Best Local Similarity 21.4%; Pred. No. 0.69; Mismatches 122; Gaps 21;  
Matches 91; Conservative 47;  
QY 8 LESVIVSRHGVRAPTKATQMLQMDVTDPAMTPWPKLGLTTPRGGLIAYLGHYWRQLV 67  
DB 33 LRFVTLVYRHGDRSPVKT----YPKDPYQEEKPQGFQLTKEGMLQHWELGQALRQRY- 87  
QY 68 ADGLLPKCGCQSQQVAIIADVD-ERTRKGEAFAGLAPDCAITVQTADTSSPDPLFN 126  
DB 88 -HGFL---NTSYHRQEVYVRSTDFDRTLMSEANLGLFPP-----NEVOHFNPNISWQ 137  
QY 127 PLKTVGCOLDNANVTDAILERAGGSIADFTGHYQTAFARELERVLNFFQSNLCLREKQDE 186  
DB 138 PI-----PVHTVPITE-----DLKLFPLGCPCEYEQQLNE 168  
QY 187 SCSTIQALPSELKVSADCVSLTGVASLASMTEIFL-----LQQAQGMPEFGWG 235  
DB 169 ---TROTPEYQNRISQNAQFLNANVANETGLTNVLTETLNWVYDTLFCQTHGLLPPWA 224  
QY 236 -----RITDSSOWN--TLLSLHNAQFDLLQRTPEVARSR-----ATPLDLIKTALT 280

```
DB 225 SPQTVRLSQLDFSLFGLFHE-----QVQKARLQGGVLLAQILKNTLMATT 274
QY 281 PPPPOKQAYGVTLPTSVLFIAGHDNTNLALGGALELNWTLPOQDNTPPGGELVFERWR 340
DB 275 SQFPK-----LLVYSHDITLVALQALNVY-----NGKQAYASCHIFELYQ- 317
QY 341 LSDNSQWIOVSLVPTQLOMRKDTPLSLNTPPGVKKLTLAGEERNAQMGCSLAGFTQIV 400
DB 318 -EDNGNF-SVEMYFRNDSK---KAPWPLILP-----GCPHR-----CPLQDFLFL- 357
QY 401 NEARIP 406
DB 358 TEVIP 363

RESULT 14
DIVL CAUCR STANDARD; PRT; 769 AA.
AC Q9RQ09; Q9A2S2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Sensor protein divL (SC 2.7.3.-).
GN DIVL OR CC3484.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
OC Caulobacteraceae; Caulobacter.
OC NCBI_TaxID=155892;
RN [1]
SEQUENCE FROM N.A., AND AUTOPHOSPHORYLATION SITE.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=20027501; PubMed=10557274;
RA Wu J., Ohta N., Zhao J.L., Newton A.;
RT "A novel bacterial tyrosine kinase essential for cell division and
RT differentiation."
RL Proc. Natl. Acad. Sci. U.S.A. 96:13068-13073 (1999).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Niernan W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J.A., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadtke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141 (2001).
CC -1- FUNCTION: Required for cell division and growth. It catalyzes the
CC phosphorylation of CtrA and activates transcription in vitro of
CC the cell cycle-regulated flf promoter.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- PTM: Autophosphorylated.
CC -1- SIMILARITY: Contains 1 histidine kinase domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC -----
CC ENBL; AF083422; AAF08344.2; -.
CC ENBL; AE006007; AAK25446.1; -.
CC PIR; B87681; B87681.
CC TIGR; CC3484; -.
CC
CC Prossite; Q9RQ09; -.
CC InterPro; IPR003594; ATPbind_ATPase.
CC InterPro; IPR004358; Bact_sens_pr_C.
CC InterPro; IPR003661; His_Kin_N.
```

```
DR InterPro; IPR005467; His_kinase.
DR Pfam; PF02518; HATPase_C; 1.
DR Pfam; PF00512; H1SKA; 1.
DR PRINTS; PRO0344; BCTRLSENSOR.
DR SMART; SMC0387; HATPase_C; 1.
DR SMART; SMC0388; H1SKA; 1.
DR PROSITE; PS01019; HIS_KIN; 1.
KW Sensory transduction; Transferrase; Kinase; Transmembrane;
KW Phosphorylation; Complete proteome.
FT TRANSMEM 6 26 POTENTIAL.
FT DOMAIN 547 758 HISTIDINE KINASE.
FT DOMAIN 9 221 ALA-RICH.
FT MOD RES 550 550 PHOSPHORYLATION (AUTO-).
FT CONFLICT 200 200 Q -> H (IN REF. 1).
FT CONFLICT 216 216 V -> E (IN REF. 1).
SQ SEQUENCE 769 AA; 82796 MW; 002B2428F18A57EF CRC64;

Query Match 4.8%; Score 103; DB 1; Length 769;
Best Local Similarity 22.9%; Pred. No. 1.8;
Matches 91; Conservative 32; Mismatches 144; Indels 130; Gaps 19;

QY 24 KATQL--MQDVTDPANPTWPKLGELT-----PQGELIAYLGHYWRQL----- 66
DB 345 KAELARYEDLGFQADDLWLPDGRILKVVRRQHPHPLGMLLIYSDITGELRLKQAQYNALI 404
QY 67 -VADGLLPKCGCPQSGQVAIADVDERTKRTGEFAA-----GLAPDCA 109
DB 405 VQQAATLDKL---NDAAVAVFGS-DGRLRLHNEAFETFWNTPHALEAAGDFGVVVELCV 459
QY 110 ITVH-----TQADTSSPDPLF-----NPLKTVGCOLDNANVTD 142
DB 460 PRLHDLJSFWRELKGRVADPDQWRAPTSGEVRTSDSRIVYQSRPLPDGATLIAFADVT 519
QY 143 AILERAGGSIADFTGHYQTAFRELERVLNFPQSNLCLKREKQDECSLTQALPSELKVA 202
DB 520 T-----RDLQSA-----ADRSAAALAE-----ERLKR 542
QY 203 DCVSLTGAVS--LASMLTEIF-----LQQAQGMPEPGWGRITSDHOWNTLLS-----LH 250
DB 543 DFV---GNVSYEILRTPLTITIGYSELLERADGTSERGRNHAARAATOLARSIDVDLD 599
QY 251 NAQFDLQRTPEVARSRATPLDLIKTALTTPHPQKQAVGVTLPTSVLFIAGHDNTLANL 310
DB 600 MAQIDAGEMALETIEDIRVS---DLLLNAQERALKDAQLGCVTLAVE-----CEEDVGLIR- 651
QY 311 GGALELNWTLPGQPDN-----TPQGELVFERWRRLSD 343
DB 652 GDGKRLAQTLDLHLVENALRQTTPGGRVTLSSARRALGE 688

RESULT 15
SECC_ECOLI STANDARD; PRT; 1048 AA.
ID SECC_ECOLI
AC P13458;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Exonuclease sbcc.
GN SBCC OR RNAA OR B0397.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12; PubMed=2530497;
RX MEDLINE=90045931; PubMed=2530497;
RA Naom I.S., Morton S.J., Leach D.R.F., Lloyd R.G.;
RT "Molecular organization of sbcc, a gene that affects genetic
RT recombination and the viability of DNA palindromes in Escherichia
RT coli K-12."
RL Nucleic Acids Res. 17:8033-8046 (1989).
RN [2]
```



SEQUENCE FROM N.A.  
STRAIN=X12 / MG1655;  
MEDLINE=97426617; PubMed=9278503;  
Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
Riley M., Collado-Vides J., Glaesner J.D., Rode C.K., Mayhew G.F.,  
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
Mau B., Shao Y.,  
"The complete genome sequence of *Escherichia coli* K-12.";  
Science 277:1453-1474 (1997).  
[3]  
SEQUENCE FROM N.A.  
Duncan M., Allen E., Araujo R., Aparicio A.M., Chung E., Davis K.,  
Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H.,  
Lin D., Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;  
Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
[4]  
SEQUENCE OF 378-1048 FROM N.A.  
MEDLINE=92078081; PubMed=1744033;  
Reeder T., Schleif R.;  
"Mapping, sequence, and apparent lack of function of *araJ*, a gene of  
the *Escherichia coli* arabinose regulon.";  
J. Bacteriol. 173:7765-7771 (1991).  
[5]  
CHARACTERIZATION.  
MEDLINE=931146416; PubMed=1490631;  
Leach D.R.F., Lloyd R.G., Coulson A.F.;  
"The SbcCD protein of *Escherichia coli* is related to two putative  
nucleases in the UvrA superfamily of nucleotide-binding proteins.";  
Genetica 87:95-100 (1992).  
[6]  
CHARACTERIZATION.  
MEDLINE=20345301; PubMed=10886369;  
Slupska M.M., Chiang J.-H., Luther W.M., Stewart J.L., Amil L.,  
Conrad A., Miller J.H.;  
"Genes involved in the determination of the rate of inversions at  
short inverted repeats.";  
Genes Cells 5:425-437 (2000).  
[7]  
CHARACTERIZATION.  
MEDLINE=98318595; PubMed=9653124;  
Connelly J.C., Kirkham L.A., Leach D.R.;  
"The SbcCD nuclease of *Escherichia coli* is a structural maintenance of  
chromosomes (SMC) family protein that cleaves hairpin DNA.";  
Proc. Natl. Acad. Sci. U.S.A. 95:7969-7974 (1998).  
[8]  
CHARACTERIZATION.  
MEDLINE=99178292; PubMed=9927737;  
Connelly J.C., de Leau E.S., Leach D.R.;  
"DNA cleavage and degradation by the SbcCD protein complex from  
*Escherichia coli*.";  
Nucleic Acids Res. 27:1039-1046 (1999).  
CC -1- FUNCTION: SbcCD cleaves DNA hairpin structures. These structures  
can inhibit DNA replication and are intermediates in certain DNA  
recombination reactions. The complex acts as a 3'-5' double  
strand exonuclease that can open hairpins. It also has a 5'  
single-strand endonuclease activity.  
CC -1- SUBUNIT: Heterodimer of sbcC and sbcD.  
CC -1- SIMILARITY: Belongs to the SMC family. SbcC subfamily.  
-----  
THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement. See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@sb-sib.ch](mailto:license@sb-sib.ch).  
-----  
EMBL; X15981; CAA34104.1; --  
DR EMBL; AE000146; AAC73500.1; --  
DR EMBL; U73857; AAB18121.1; --  
DR EMBL; M64787; AAB23473.1; --  
DR PIR; J50350; BVESCC.  
DR EcoGene; EG10927; sbcC.

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 12:34:35 ; Search time 37.1429 Seconds  
(without alignments)  
3482.836 Million cell updates/sec

Title: US-10-021-723B-13  
Perfect score: 2153  
Sequence: 1 QSEPELKESVIVSRHGVR.....CSLAGFTQIVNEARIPACSL 410

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL 25:\*

- 1: sp\_archea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2073	96.3	432	2 Q8GN88	Q8GN88 escherichia
2	2071	96.2	432	2 Q8KDD6	Q8KDD6 escherichia
3	2071	95.2	432	2 Q8KDD7	Q8KDD7 escherichia
4	2071	96.2	432	2 Q8KDD8	Q8KDD8 escherichia
5	2071	96.2	432	2 Q8KDD9	Q8KDD9 escherichia
6	2070	96.1	432	2 Q8KDD5	Q8KDD5 escherichia
7	2067	96.0	432	2 Q8KDD3	Q8KDD3 shigella fl
8	2044	94.9	442	16 Q83RW2	Q83RW2 shigella fl
9	2042	94.8	434	16 Q8KCD9	Q8KCD9 escherichia
10	2042	94.8	446	16 Q8CW75	Q8CW75 escherichia
11	2040	94.8	432	16 Q7UD08	Q7UD08 shigella fl
12	954	44.3	441	16 Q8ZFP6	Q8ZFP6 vercinia de
13	551	25.6	413	16 Q8XB26	Q8XB26 escherichia
14	543	25.2	413	16 Q8Z7F1	Q8Z7F1 salmonella
15	543	25.2	413	16 Q7UD02	Q7UD02 shigella fl
16	543	25.2	421	16 Q83RV6	Q83RV6 shigella fl

17	534.5	24.8	414	16	Q9AAQ4	Q9AAQ4 caulobacter
18	511	23.7	392	16	Q8PP76	Q8PP76 xanthomonas
19	497	23.1	443	16	Q8P330	Q8P330 xanthomonas
20	455.5	21.2	435	16	Q8PF53	Q8PF53 xanthomonas
21	432	20.1	428	2	Q8GD20	Q8GD20 pseudomonas
22	394	18.3	318	16	Q8CW72	Q8CW72 escherichia
23	357.5	16.6	414	16	Q8PPX6	Q8PPX6 xanthomonas
24	356	16.5	421	2	Q84CN9	Q84CN9 klebsiella
25	348	16.2	421	2	Q8WSY1	Q8WSY1 klebsiella
26	227.5	10.6	419	2	Q8VQS2	Q8VQS2 klebsiella
27	184	8.5	426	4	Q8BZG2	Q8BZG2 homo sapien
28	133	6.2	416	5	Q19390	Q19390 caenorhabdi
29	132.5	6.2	531	5	Q9GNZ3	Q9GNZ3 leishmania
30	123.5	5.7	333	4	Q8BZG3	Q8BZG3 homo sapien
31	123.5	5.7	395	5	Q9VD68	Q9VD68 drosophila
32	123.5	5.7	542	5	Q9UIA2	Q9UIA2 leishmania
33	122.5	5.7	683	5	O00838	O00838 leishmania
34	122.5	5.7	707	5	O00839	O00839 leishmania
35	121.5	5.6	447	5	Q9USU3	Q9USU3 drosophila
36	120.5	5.6	447	5	Q9USV1	Q9USV1 drosophila
37	118.5	5.5	442	3	Q8WZJ5	Q8WZJ5 aspergillus
38	118.5	5.5	447	5	O00092	O00092 aspergillus
39	117.5	5.5	447	5	Q9USU5	Q9USU5 drosophila
40	117.5	5.5	447	5	Q9TW17	Q9TW17 drosophila
41	117.5	5.5	447	5	Q9TW19	Q9TW19 drosophila
42	117.5	5.5	447	5	Q9U5T5	Q9U5T5 drosophila
43	117.5	5.5	447	5	Q9TW53	Q9TW53 drosophila
44	117.5	5.5	447	5	Q9USU2	Q9USU2 drosophila
45	117.5	5.5	447	5	O97187	O97187 drosophila

ALIGNMENTS

RESULT 1

Q8GN88 PRELIMINARY; PRT; 432 AA.  
AC Q8GN88;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE APPA.  
GN APPA.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
RX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Zhang Z., He J., Yao B., Zhou Y., Chen Y., Yi Y.;  
RT "Production of phytase and acid phosphatase by use of silkworm-  
RT bioreactor";  
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
[2]  
RP SEQUENCE FROM N.A.  
RA Chen Y., Zhu Z., Zhang Z., He J.;  
RT "Cloning and overexpression of phytase gene appA from Escherichia  
RT coli";  
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
EMBL; AF537219; AAN28334.1; -;  
DR GO; GO:0003993; F:acid phosphatase activity; IEA.  
DR InterPro; IPR000560; HisAc\_Phsptase.  
DR Pfam; PF00328; acid phosphat; 1  
DR PROSITE; PS00616; HIS-ACID-PHOSPHAT\_1; 1.  
DR PROSITE; PS00778; HIS-ACID-PHOSPHAT\_2; 1.  
SQ SEQUENCE 432 AA; 47055 MW; 5B355D76E7377737 CRC64;

Query Match 96.3%; Score 2073; DB 2; Length 432;  
Best Local Similarity 97.8%; Pred. No. 2.2e-165;  
Matches 401; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

Qy 1 QSEPELKESVIVSRHGVRAPKATQLQMDVTPDAWTPVVKLGELTPRGGLIAYLGH 60  
|||||

Db 23 QSEPELKLESVIVSRHGVRAPTKATQMQDVTPDAMPVTPVKLGWLTTPRGGLIAYLGH 82  
QY 61 YWRQRLVADGILLPKCCPGSQGQVAIIADVDERTKRTGEAFAAGLAPDCAITVHTQADTSS 120  
Db 83 YQORQLVADGILLAKKCGPQSGQVAIIADVDERTKRTGEAFAAGLAPDCAITVHTQADTSS 142  
QY 121 PDPLFNLKTVGVCOLDNANVTDAILERAGGSIADFTGHYQTAFARELERVLPQSNLCLK 180  
Db 143 PDPLFNLKTVGVCOLDNANVTDAILERAGGSIADFTGHYQTAFARELERVLPQSNLCLK 202  
QY 181 REKQDESCSLTOALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 240  
Db 203 REKQDESCSLTOALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 262  
QY 241 HONWTLISLHNAQFDLLQRTPEVARSRATPLLDLIKTALTPHPQKQAYGVLTPTSVLFI 300  
Db 263 HONWTLISLHNAQFYLLQRTPEVARSRATPLLDLIKTALTPHPQKQAYGVLTPTSVLFI 322  
QY 301 AGHDTNLANLGGALELNTLPGQPDNTPPGGELVFERWRRLSDNSQWIVQSLVFTLQOM 360  
Db 323 AGHDTNLANLGGALELNTLPGQPDNTPPGGELVFERWRRLSDNSQWIVQSLVFTLQOM 382  
QY 361 RDKTPLSLNTPPGSVKLTLAGCERNAQGMCSLAGFTQIVNEARIPACSL 410  
Db 383 RDKTPLSLNTPPGSVKLTLAGCERNAQGMCSLAGFTQIVNEARIPACSL 432

## RESULT 2

Q8RKD6 PRELIMINARY; PRT; 432 AA.  
AC Q8RKD6; 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, last annotation update)  
DE Periplasmic phosphoanhydride phosphohydrolase.  
GN APPA.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90368616; PubMed=2168385;  
RA Dassa J., Marck C., Boquet P.-L.L.;  
RT "The complete nucleotide sequence of the Escherichia coli gene appA  
RT reveals significant homology between pH 2.5 acid phosphatase and  
RT glucose-1-phosphatase.";  
RL J. Bacteriol. 172:5497-5500(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93054596; PubMed=1429631;  
RA Ostanin K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.M.,  
RA Van Etten R.L.;  
RT "Overexpression, site-directed mutagenesis, and mechanism of  
RT Escherichia coli acid phosphatase.";  
RL J. Biol. Chem. 267:22830-22836(1992).  
DR EMBL; L03374; AAA0006.1;  
DR GO; GO:0003993; F:acid phosphatase activity; IEA.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR000560; HisAc\_phspatse.  
DR Pfam; PF00328; acid\_phosphat; 1.  
DR PROSITE; PS00616; HIS\_ACID\_PHOSPHAT\_1; 1.  
DR PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.  
KW Hydrolase.  
SQ SEQUENCE 432 AA; 46971 MW; AF86C41EA6193AC5 CRC64;

Query Match 96.2%; Score 2071; DB 2; Length 432;  
Best Local Similarity 97.8%; Pred. No. 3.2e-165;  
Matches 401; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
QY 1 QSEPELKLESVIVSRHGVRAPTKATQMQDVTPDAMPVTPVKLGWLTTPRGGLIAYLGH 60  
Db 23 QSEPELKLESVIVSRHGVRAPTKATQMQDVTPDAMPVTPVKLGWLTTPRGGLIAYLGH 82

QY 61 YWRQRLVADGILLPKCCPGSQGQVAIIADVDERTKRTGEAFAAGLAPDCAITVHTQADTSS 120  
Db 83 YQORQLVADGILLAKKCGPQSGQVAIIADVDERTKRTGEAFAAGLAPDCAITVHTQADTSS 142  
QY 121 PDPLFNLKTVGVCOLDNANVTDAILERAGGSIADFTGHYQTAFARELERVLPQSNLCLK 180  
Db 143 PDPLFNLKTVGVCOLDNANVTDAILERAGGSIADFTGHYQTAFARELERVLPQSNLCLK 202  
QY 181 REKQDESCSLTOALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 240  
Db 203 REKQDESCSLTOALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 262  
QY 241 HONWTLISLHNAQFDLLQRTPEVARSRATPLLDLIKTALTPHPQKQAYGVLTPTSVLFI 300  
Db 263 HONWTLISLHNAQFYLLQRTPEVARSRATPLLDLIKTALTPHPQKQAYGVLTPTSVLFI 322  
QY 301 AGHDTNLANLGGALELNTLPGQPDNTPPGGELVFERWRRLSDNSQWIVQSLVFTLQOM 360  
Db 323 AGHDTNLANLGGALELNTLPGQPDNTPPGGELVFERWRRLSDNSQWIVQSLVFTLQOM 382  
QY 361 RDKTPLSLNTPPGSVKLTLAGCERNAQGMCSLAGFTQIVNEARIPACSL 410  
Db 383 RDKTPLSLNTPPGSVKLTLAGCERNAQGMCSLAGFTQIVNEARIPACSL 432

## RESULT 3

Q8RKD7 PRELIMINARY; PRT; 432 AA.  
AC Q8RKD7; 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, last annotation update)  
DE Periplasmic phosphoanhydride phosphohydrolase.  
GN APPA.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90368616; PubMed=2168385;  
RA Dassa J., Marck C., Boquet P.-L.L.;  
RT "The complete nucleotide sequence of the Escherichia coli gene appA  
RT reveals significant homology between pH 2.5 acid phosphatase and  
RT glucose-1-phosphatase.";  
RL J. Bacteriol. 172:5497-5500(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93054596; PubMed=1429631;  
RA Ostanin K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.M.,  
RA Van Etten R.L.;  
RT "Overexpression, site-directed mutagenesis, and mechanism of  
RT Escherichia coli acid phosphatase.";  
RL J. Biol. Chem. 267:22830-22836(1992).  
DR EMBL; L03373; AAA0005.1;  
DR GO; GO:0003993; F:acid phosphatase activity; IEA.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR000560; HisAc\_phspatse.  
DR Pfam; PF00328; acid\_phosphat; 1.  
DR PROSITE; PS00616; HIS\_ACID\_PHOSPHAT\_1; 1.  
DR PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.  
KW Hydrolase.  
SQ SEQUENCE 432 AA; 46971 MW; 755D5E4B1AD916A6 CRC64;

Query Match 96.2%; Score 2071; DB 2; Length 432;  
Best Local Similarity 97.8%; Pred. No. 3.2e-165;  
Matches 401; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
QY 1 QSEPELKLESVIVSRHGVRAPTKATQMQDVTPDAMPVTPVKLGWLTTPRGGLIAYLGH 60  
Db 23 QSEPELKLESVIVSRHGVRAPTKATQMQDVTPDAMPVTPVKLGWLTTPRGGLIAYLGH 82

QY 61 YWQRLVADGLLPKKCGPQSGQVAIIADVDERTKTGEAFAAGLAPDCAITVHTQADTSS 120  
DB 83 YQORLVADGLLAKKCGPQSGQVAIIADVDERTKTGEAFAAGLAPDCAITVHTQADTSS 142  
QY 121 PDPLFNLKTVGCQLDNANVTDAIIRAGGSIADFTGHYQTAFRELERVLPQSNLCLK 180  
DB 143 PDPLFNLKTVGCQLDNANVTDAIIRAGGSIADFTGHYQTAFRELERVLPQSNLCLK 202  
QY 181 REKQDESCSLTOALPSELKVSADCVSLTGVASLASMTEIFLLQOAGMPGPGWGRITDS 240  
DB 203 REKQDESCSLTOALPSELKVSADCVSLTGVASLASMTEIFLLQOAGMPGPGWGRITDS 262  
QY 241 HOWNTLISLHNAQFDLQRTPEVARSRATPLLDLTKTALTPHPPOKQAYGVLTPTSVLFI 300  
DB 263 HOWNTLISLHNAQFYLLQRTPEVARSRATPLLDLTKTALTPHPPOKQAYGVLTPTSVLFI 322  
QY 301 AGHDTNLANLGGALELNLWTLPGQDNTPPGGELVFERWRRLSDNSQWISLVFOTLQOM 360  
DB 323 AGHDTNLANLGGALELNLWTLPGQDNTPPGGELVFERWRRLSDNSQWISLVFOTLQOM 382  
QY 361 RDKTPLSLNTPPGGVKLTLAGCERNAGCMCSLAGFTQIVNEARIPACSL 410  
DB 383 RDKTPLSLNTPPGGVKLTLAGCERNAGCMCSLAGFTQIVNEARIPACSL 432

## RESULT 4

QSRKEO PRELIMINARY; PRT; 432 AA.  
AC QSRKEO;  
DT 01-JUN-2002 (TREMELrel. 21, Created)  
DT 01-JUN-2002 (TREMELrel. 21, Last sequence update)  
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)  
DE Periplasmic phosphoanhydride phosphohydrolase.  
GN APPA.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90368616; PubMed=2168385;  
RA Dassa J., Marck C., Boquet P.-L.L.;  
RT "The complete nucleotide sequence of the Escherichia coli gene appA  
RT reveals significant homology between pH 2.5 acid phosphatase and  
RT glucose-1-phosphatase.";  
RL J. Bacteriol. 172:5497-5500(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93054596; PubMed=1429631;  
RA Ostanin K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.M.,  
RA Van Etten R.L.;  
RT "Overexpression, site-directed mutagenesis, and mechanism of  
RT Escherichia coli acid phosphatase.";  
RL J. Biol. Chem. 267:22830-22836(1992).  
DR EMBL; L03372; AAA00004.1; -;  
DR GO; GO:0003993; F:acid phosphatase activity; IEA.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR000560; HisAc\_phosphatse.  
DR Pfam; PF00328; acid\_phosphat; 1.  
DR PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.  
KW Hydrolase.  
SQ SEQUENCE 432 AA; 46971 MW; 9A85536B57FCCB5 CRC64;

Query Match 96.2%; Score 2071; DB 2; Length 432;  
Best Local Similarity 97.8%; Pred. No. 3.2e-165;  
Matches 401; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
QY 1 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTDPDWFTWPKVGLTPRGGLIAYLGH 60  
DB 23 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTDPDWFTWPKVGLTPRGGLIAYLGH 82  
QY 61 YWQRLVADGLLPKKCGPQSGQVAIIADVDERTKTGEAFAAGLAPDCAITVHTQADTSS 120

DB 83 YQORLVADGLLAKKCGPQSGQVAIIADVDERTKTGEAFAAGLAPDCAITVHTQADTSS 142  
QY 121 PDPLFNLKTVGCQLDNANVTDAIIRAGGSIADFTGHYQTAFRELERVLPQSNLCLK 180  
DB 143 PDPLFNLKTVGCQLDNANVTDAIIRAGGSIADFTGHYQTAFRELERVLPQSNLCLK 202  
QY 181 REKQDESCSLTOALPSELKVSADCVSLTGVASLASMTEIFLLQOAGMPGPGWGRITDS 240  
DB 203 REKQDESCSLTOALPSELKVSADCVSLTGVASLASMTEIFLLQOAGMPGPGWGRITDS 262  
QY 241 HOWNTLISLHNAQFDLQRTPEVARSRATPLLDLTKTALTPHPPOKQAYGVLTPTSVLFI 300  
DB 263 HOWNTLISLHNAQFYLLQRTPEVARSRATPLLDLTKTALTPHPPOKQAYGVLTPTSVLFI 322  
QY 301 AGHDTNLANLGGALELNLWTLPGQDNTPPGGELVFERWRRLSDNSQWISLVFOTLQOM 360  
DB 323 AGHDTNLANLGGALELNLWTLPGQDNTPPGGELVFERWRRLSDNSQWISLVFOTLQOM 382  
QY 361 RDKTPLSLNTPPGGVKLTLAGCERNAGCMCSLAGFTQIVNEARIPACSL 410  
DB 383 RDKTPLSLNTPPGGVKLTLAGCERNAGCMCSLAGFTQIVNEARIPACSL 432

## RESULT 5

QSRKEO PRELIMINARY; PRT; 432 AA.  
AC QSRKEO;  
DT 01-JUN-2002 (TREMELrel. 21, Created)  
DT 01-JUN-2002 (TREMELrel. 21, Last sequence update)  
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)  
DE Periplasmic phosphoanhydride phosphohydrolase.  
GN APPA.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90368616; PubMed=2168385;  
RA Dassa J., Marck C., Boquet P.-L.L.;  
RT "The complete nucleotide sequence of the Escherichia coli gene appA  
RT reveals significant homology between pH 2.5 acid phosphatase and  
RT glucose-1-phosphatase.";  
RL J. Bacteriol. 172:5497-5500(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93054596; PubMed=1429631;  
RA Ostanin K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.M.,  
RA Van Etten R.L.;  
RT "Overexpression, site-directed mutagenesis, and mechanism of  
RT Escherichia coli acid phosphatase.";  
RL J. Biol. Chem. 267:22830-22836(1992).  
DR EMBL; L03370; AAA00002.1; -;  
DR GO; GO:0003993; F:acid phosphatase activity; IEA.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR000560; HisAc\_phosphatse.  
DR Pfam; PF00328; acid\_phosphat; 1.  
DR PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.  
KW Hydrolase.  
SQ SEQUENCE 432 AA; 46971 MW; 5BBD632D4682EABF CRC64;

Query Match 96.2%; Score 2071; DB 2; Length 432;  
Best Local Similarity 97.8%; Pred. No. 3.2e-165;  
Matches 401; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
QY 1 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTDPDWFTWPKVGLTPRGGLIAYLGH 60  
DB 23 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTDPDWFTWPKVGLTPRGGLIAYLGH 82  
QY 61 YWQRLVADGLLPKKCGPQSGQVAIIADVDERTKTGEAFAAGLAPDCAITVHTQADTSS 120  
DB 83 YQORLVADGLLAKKCGPQSGQVAIIADVDERTKTGEAFAAGLAPDCAITVHTQADTSS 142

QY 121 PDPLFNLKTVGCQOLDNANVTDAILRAGGSIADFTGHTQATAFRELERVLPFQSNLCLK 180  
 DB 143 PDPLFNLKTVGCQOLDNANVTDAILRAGGSIADFTGHTQATAFRELERVLPFQSNLCLK 202  
 QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGNPEPGWGRITDS 240  
 DB 203 REKQDESCSLTQALPSELKVSADNVSITGAVSLASMLTEIFLLQQAQGNPEPGWGRITDS 262  
 QY 241 HONWTLISLHNAQFDLLQRTPEVARSRATPLDLLIKTALTTPHPKQAYGVTLPTSVLFI 300  
 DB 263 HONWTLISLHNAQFYLLQRTPEVARSRATPLDLLIKTALTTPHPKQAYGVTLPTSVLFI 322  
 QY 301 AGHDTNLANLGGALELNLWTLPGQPDNTPPGGELVFERWRRLSDNSQWIVQSVLFTQLQOM 360  
 DB 323 AGHDTNLANLGGALELNLWTLPGQPDNTPPGGELVFERWRRLSDNSQWIVQSVLFTQLQOM 382  
 QY 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQMCISLAGFTQVNEARIPACSL 410  
 DB 383 RDKTPLSLNTPPGEVKLTLAGCEERNAQMCISLAGFTQVNEARIPACSL 432

## RESULT 6

QSRKD9 PRELIMINARY; PRT; 432 AA.  
 AC QSRKD9;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Periplasmic phosphoanhydride phosphohydrolase.  
 GN APPA.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90368616; PubMed=2168385;  
 RA Dassa J., Marck C., Boquet P.-L.L.;  
 RT "The complete nucleotide sequence of the Escherichia coli gene appA  
 RT reveals significant homology between pH 2.5 acid phosphatase and  
 RT glucose-1-phosphatase."  
 RL J. Bacteriol. 172:5497-5500 (1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93054596; PubMed=1429631;  
 RA Ostanin K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.M.,  
 RA Van Erten R.L.;  
 RT "Overexpression, site-directed mutagenesis, and mechanism of  
 RT Escherichia coli acid phosphatase."  
 RL J. Biol. Chem. 267:22830-22836 (1992).  
 DR EMBL; L03371; AAA00003.1; -  
 DR GO; GO:0003993; F:acid phosphatase activity; IEA.  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR InterPro; IPR000560; HisAc phsphtse.  
 DR Pfam; PF00328; acid\_phosphat; 1.  
 DR PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.  
 KW Hydrolase.  
 SQ SEQUENCE 432 AA; 47033 MW; 9F2959DF9C368175 CRC64;

Query Match 96.1%; Score 2070; DB 2; Length 432;

Best Local Similarity 97.8%; Pred. No. 3.9e-165;

Matches 401; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 1 QSEPELKLESVIVSRHGVRAPTKATQIMQDVTDPDAPWTPVKLGELTPRGGLIAYLGH 60  
 DB 23 QSEPELKLESVIVSRHGVRAPTKATQIMQDVTDPDAPWTPVKLGELTPRGGLIAYLGH 82  
 QY 61 YWRQLVADGLLKKCGCPQSGQVAIIADVDERTKTGEAFAGLAPDCAITVHTQADTSS 120  
 DB 83 YWRQLVADGLLKKCGCPQSGQVAIIADVDERTKTGEAFAGLAPDCAITVHTQADTSS 142  
 QY 121 PDPLFNLKTVGCQOLDNANVTDAILRAGGSIADFTGHTQATAFRELERVLPFQSNLCLK 180

DB 143 PDPLFNLKTVGCQOLDNANVTDAILRAGGSIADFTGHTQATAFRELERVLPFQSNLCLK 202  
 QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGNPEPGWGRITDS 240  
 DB 203 REKQDESCSLTQALPSELKVSADNVSITGAVSLASMLTEIFLLQQAQGNPEPGWGRITDS 262  
 QY 241 HONWTLISLHNAQFDLLQRTPEVARSRATPLDLLIKTALTTPHPKQAYGVTLPTSVLFI 300  
 DB 263 HONWTLISLHNAQFYLLQRTPEVARSRATPLDLLIKTALTTPHPKQAYGVTLPTSVLFI 322  
 QY 301 AGHDTNLANLGGALELNLWTLPGQPDNTPPGGELVFERWRRLSDNSQWIVQSVLFTQLQOM 360  
 DB 323 AGHDTNLANLGGALELNLWTLPGQPDNTPPGGELVFERWRRLSDNSQWIVQSVLFTQLQOM 382  
 QY 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQMCISLAGFTQVNEARIPACSL 410  
 DB 383 RDKTPLSLNTPPGEVKLTLAGCEERNAQMCISLAGFTQVNEARIPACSL 432

## RESULT 7

QSRKD5 PRELIMINARY; PRT; 432 AA.  
 AC QSRKD5;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Periplasmic phosphoanhydride phosphohydrolase.  
 GN APPA.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90368616; PubMed=2168385;  
 RA Dassa J., Marck C., Boquet P.-L.L.;  
 RT "The complete nucleotide sequence of the Escherichia coli gene appA  
 RT reveals significant homology between pH 2.5 acid phosphatase and  
 RT glucose-1-phosphatase."  
 RL J. Bacteriol. 172:5497-5500 (1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93054596; PubMed=1429631;  
 RA Ostanin K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.M.,  
 RA Van Erten R.L.;  
 RT "Overexpression, site-directed mutagenesis, and mechanism of  
 RT Escherichia coli acid phosphatase."  
 RL J. Biol. Chem. 267:22830-22836 (1992).  
 DR EMBL; L03371; AAA00003.1; -  
 DR GO; GO:0003993; F:acid phosphatase activity; IEA.  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR InterPro; IPR000560; HisAc phsphtse.  
 DR Pfam; PF00328; acid\_phosphat; 1.  
 DR PROSITE; PS00616; HIS\_ACID\_PHOSPHAT\_1; 1.  
 KW Hydrolase.  
 SQ SEQUENCE 432 AA; 46990 MW; 951F393EA9A1A47C CRC64;

Query Match 96.0%; Score 2067; DB 2; Length 432;

Best Local Similarity 97.8%; Pred. No. 6.9e-165;

Matches 401; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 QSEPELKLESVIVSRHGVRAPTKATQIMQDVTDPDAPWTPVKLGELTPRGGLIAYLGH 60  
 DB 23 QSEPELKLESVIVSRHGVRAPTKATQIMQDVTDPDAPWTPVKLGELTPRGGLIAYLGH 82  
 QY 61 YWRQLVADGLLKKCGCPQSGQVAIIADVDERTKTGEAFAGLAPDCAITVHTQADTSS 120  
 DB 83 YWRQLVADGLLKKCGCPQSGQVAIIADVDERTKTGEAFAGLAPDCAITVHTQADTSS 142  
 QY 121 PDPLFNLKTVGCQOLDNANVTDAILRAGGSIADFTGHTQATAFRELERVLPFQSNLCLK 180  
 DB 143 PDPLFNLKTVGCQOLDNANVTDAILRAGGSIADFTGHTQATAFRELERVLPFQSNLCLK 202

```
QY 181 REKQDESCSLTOALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMGPEPGWGRITDS 240
Db 203 REKQDESCSLTOALPSELKVSADNVSLTGAVSLASMLTEIFLLQQAQMGPEPGWGRITDS 262
QY 241 HQWNTLLSLHNAQFDLIQRTPEVARSRATPLDLIKTALTALTPHPPOKQAYGVTLPTSVLFI 300
Db 263 HQWNTLLSLHNAQFDLIQRTPEVARSRATPLDLIKTALTALTPHPPOKQAYGVTLPTSVLFI 322
QY 301 AGHDTNLANLGGALELNTLPGQPDNTPGCGELVFERWRRLSDNSOMIQVSLVFQTLQOM 360
Db 323 AGADTNLANLGGALELNTLPGQPDNTPGCGELVFERWRRLSDNSOMIQVSLVFQTLQOM 382
QY 361 RDKTPLSLNTPPGVEVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 410
Db 383 RDKTPLSLNTPPGVEVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 432

RESULT 8
Q83RW2 PRELIMINARY; PRT; 442 AA.
ID AC Q83RW2;
AC Q83RW2;
DT 01-JUN-2003 (TREMREL. 24, Created)
DT 01-JUN-2003 (TREMREL. 24, Last sequence update)
DT 01-OCT-2003 (TREMREL. 25, Last annotation update)
DE Phosphoanhydride phosphorylase, pH 2.5 acid phosphatase,
DE periplasmic.
GN APPA OR SF0982.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qi D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Chang H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157."
RL Nucleic Acids Res. 30:4432-4441 (2002).
DR EMBL; AE015127; AAA2610.1; -.
DR GO; GO:0003993; F:acid phosphatase activity; IEA.
DR InterPro; IPR000560; HisAc_phsphtse.
DR Pfam; PF00328; acid_phosphat; 1.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
KW Complete proteome.
SQ SEQUENCE 442 AA; 48183 MW; 20BCE2C454254680 CRC64;

Query Match 94.9%; Score 2044; DB 16; Length 442;
Best Local Similarity 96.6%; Pred. No. 6.1e-163;
Matches 396; Conservative 1; Mismatches 13; Indels 0; Gaps 0;

QY 1 QSEPELKESVIVSRHGVRAPTKATQLMQDVTDPDAPWTPVKLGELTPRGGLIAYLGH 60
Db 33 QSEPELKESVIVSRHGVRAPTKATQLMQDVTDPDAPWTPVKLGELTPRGGLIAYLGH 92
QY 61 YWRQLVADGLLPKCGCPQSQGVAIIADVDERTRKTGEAFAGLAPDCAITVHTQADTSS 120
Db 93 YWRQLVADGLLPKCGCPQSQGVAIIADVDERTRKTGEAFAGLAPDCAITVHTQADTSS 152
QY 121 PDLFNLKTVGVCQLDNANVTDAILERAGGSTADFTGHVQTAFRELERLVNFPQSNCLCK 180
Db 153 PDLFNLKTVGVCQLDNANVTDAILERAGGSTADFTGHVQTAFRELERLVNFPQSNCLCK 212
QY 181 REKQDESCSLTOALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMGPEPGWGRITDS 240
Db 213 REKQDESCSLTOALPSELKVSADNVSLTGAVSLASMLTEIFLLQQAQMGPEPGWGRITDS 272
QY 241 HQWNTLLSLHNAQFDLIQRTPEVARSRATPLDLIKTALTALTPHPPOKQAYGVTLPTSVLFI 300
```

```
Db 273 HQWNTLLSLHNAQFDLIQRTPEVARSRATPLDLIKTALTALTPHPPOKQAYGVTLPTSVLFI 332
QY 301 AGHDTNLANLGGALELNTLPGQPDNTPGCGELVFERWRRLSDNSOMIQVSLVFQTLQOM 360
Db 333 AGHDTNLANLGGALELNTLPGQPDNTPGCGELVFERWRRLSDNSOMIQVSLVFQTLQOM 392
QY 361 RDKTPLSLNTPPGVEVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 410
Db 393 RDKTPLSLNTPPGVEVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 442

RESULT 9
Q8XC29 PRELIMINARY; PRT; 434 AA.
ID AC Q8XC29;
AC Q8XC29;
DT 01-MAR-2002 (TREMREL. 20, Created)
DT 01-MAR-2002 (TREMREL. 20, Last sequence update)
DT 01-OCT-2003 (TREMREL. 25, Last annotation update)
DE Phosphoanhydride phosphorylase, pH 2.5 acid phosphatase,
DE periplasmic.
GN APPA OR Z1397 OR ECS1136.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayaishi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Ikeda T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22 (2001).
DR EMBL; AE005292; AAG55528.1; ALT INIT.
DR EMBL; AP002554; BAB34559.1; -.
DR PIR; D85633; D85633.
DR PIR; H90770; H90770.
DR GO; GO:0003993; F:acid phosphatase activity; IEA.
DR InterPro; IPR000560; HisAc_phsphtse.
DR Pfam; PF00328; acid_phosphat; 1.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
KW Complete proteome.
SQ SEQUENCE 434 AA; 47337 MW; F197DF7D1869F9C4 CRC64;

Query Match 94.8%; Score 2042; DB 16; Length 434;
Best Local Similarity 96.3%; Pred. No. 8.7e-163;
Matches 395; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

QY 1 QSEPELKESVIVSRHGVRAPTKATQLMQDVTDPDAPWTPVKLGELTPRGGLIAYLGH 60
Db 25 EPEPELKESVIVSRHGVRAPTKATQLMQDVTDPDAPWTPVKLGELTPRGGLIAYLGH 84
QY 61 YWRQLVADGLLPKCGCPQSQGVAIIADVDERTRKTGEAFAGLAPDCAITVHTQADTSS 120
Db 85 YWRQLVADGLLPKCGCPQSQGVAIIADVDERTRKTGEAFAGLAPDCAITVHTQADTSS 144
```

```

QY 121 PDPLFNPLKTVGCQDQDANVTDAIILRAGGSIADFTGHYQTAFARELERVLPQSNLCIK 180
DB 145 PDPLFNPLKTVGCQDQDANVTDAIILRAGGSIADFTGHYQTAFARELERVLPQSNLCIN 204
QY 181 REKQDESCSLTOALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 240
DB 205 REKQDESCSLTOALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 264
QY 241 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLLDLIKTAFTALTPHPPOKQAYGVTLPSTVLF 300
DB 265 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLLDLIKTAFTALTPHPPOKQAYGVTLPSTVLF 324
QY 301 AGHDTNLANLGGALELNLWTLPGQDNTPPGGELVFERWRRLSDNSOWIQVSLVFQTLQOM 360
DB 325 AGHDTNLANLGGALELNLWTLPGQDNTPPGGELVFERWRRLSDNSOWIQVSLVFQTLQOM 384
QY 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIIPACSL 410
DB 385 RDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIIPACSL 434
RESULT 10
Q8CW75 PRELIMINARY; PRT; 446 AA.
ID Q8CW75
AC Q8CW75;
DT 01-MAR-2003 (TremBrel. 23, Created)
DT 01-MAR-2003 (TremBrel. 23, Last sequence update)
DE Periplasmic appA protein precursor.
GN APPA OR C1121.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Raeko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.; Perna N.T.,
RT "Extensive mosaic structure revealed by the complete genome sequence
of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
DR EMBL; A616759; AAN79589.1; -.
GO; GO:0003993; F:acid phosphatase activity; IEA.
DR InterPro; IPR000560; HisAcPhsphtase.
DR Pfam; PF00328; acid_phosphat_1
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
KW Complete proteome.
SQ SEQUENCE 446 AA; 48605 MW; F1308CD691DB9F49 CRC64;
Query Match 94.8%; Score 2042; DB 16; Length 446;
Best Local Similarity 96.3%; Pred. No. 9.1e-163;
Matches 395; Conservative 2; Mismatches 13; Indels 0; Gaps 0;
QY 1 QSEPELKESVIVSRHGVRAPTKATQLMQDVTDPDAMPTWPKLGELTPRGGLIAYLGH 60
DB 33 QSEPELKESVIVSRHGVRAPTKATQLMQDVTDPDAMPTWPKLGELTPRGGLIAYLGH 92
QY 61 YNRQRLVADGLPKCGCPSQGOVAILIADVDETRKTGEAPAGLAPDCAITVHTQADTSS 120
DB 93 YNRQRLVADGLTKKCGCPQGOVAILISVDDETRKTGEAPAGLAPDCAITVHTQADTSS 152
QY 121 PDPLFNPLKTVGCQDQDANVTDAIILRAGGSIADFTGHYQTAFARELERVLPQSNLCIK 180
DB 153 PDPLFNPLKTVGCQDQDANVTDAIILRAGGSIADFTGHYQTAFARELERVLPQSNLCIN 212
QY 181 REKQDESCSLTOALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 240
DB 213 REKQDESCSLTOALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 272

```

```

QY 241 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLLDLIKTAFTALTPHPPOKQAYGVTLPSTVLF 300
DB 273 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLLDLIKTAFTALTPHPPOKQAYGVTLPSTVLF 332
QY 301 AGHDTNLANLGGALELNLWTLPGQDNTPPGGELVFERWRRLSDNSOWIQVSLVFQTLQOM 360
DB 333 AGHDTNLANLGGALELNLWTLPGQDNTPPGGELVFERWRRLSDNSOWIQVSLVFQTLQOM 392
QY 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIIPACSL 410
DB 393 RDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIIPACAL 442
RESULT 11
Q7UD08 PRELIMINARY; PRT; 432 AA.
ID Q7UD08
AC Q7UD08;
DT 01-OCT-2003 (TremBrel. 25, Created)
DT 01-OCT-2003 (TremBrel. 25, Last sequence update)
DT 01-OCT-2003 (TremBrel. 25, Last annotation update)
DE Phosphoanhydride phosphoriase; pH 2.5 acid phosphatase.
GN APPA OR S1048.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
flexneri serotype 2a strain 2457T.";
RL Infect. Immun. 71:2775-2786 (2003).
DR EMBL; AB016981; AAP16495.1; -.
SQ SEQUENCE 432 AA; 47136 MW; 200442239BC66DFC CRC64;
Query Match 94.8%; Score 2040; DB 16; Length 432;
Best Local Similarity 96.3%; Pred. No. 1.3e-162;
Matches 395; Conservative 1; Mismatches 14; Indels 0; Gaps 0;
QY 1 QSEPELKESVIVSRHGVRAPTKATQLMQDVTDPDAMPTWPKLGELTPRGGLIAYLGH 60
DB 23 QSEPELKESVIVSRHGVRAPTKATQLMQDVTDPDAMPTWPKLGELTPRGGLIAYLGH 82
QY 61 YNRQRLVADGLPKCGCPSQGOVAILIADVDETRKTGEAPAGLAPDCAITVHTQADTSS 120
DB 83 YNRQRLVADGLLAKKGCPSQGOVAILIADVDETRKTGEAPAGLAPDCAITVHTQADTSS 142
QY 121 PDPLFNPLKTVGCQDQDANVTDAIILRAGGSIADFTGHYQTAFARELERVLPQSNLCIK 180
DB 143 PDPLFNPLKTVGCQDQDANVTDAIILRAGGSIADFTGHYQTAFARELERVLPQSNLCIN 202
QY 181 REKQDESCSLTOALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 240
DB 203 REKQDESCSLTOALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 262
QY 241 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLLDLIKTAFTALTPHPPOKQAYGVTLPSTVLF 300
DB 263 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLLDLIKTAFTALTPHPPOKQAYGVTLPSTVLF 322
QY 301 AGHDTNLANLGGALELNLWTLPGQDNTPPGGELVFERWRRLSDNSOWIQVSLVFQTLQOM 360
DB 323 AGHDTNLANLGGALELNLWTLPGQDNTPPGGELVFERWRRLSDNSOWIQVSLVFQTLQOM 382
QY 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIIPACSL 410
DB 383 RDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIIPACSL 432

```



RESULT 12  
Q8ZFP6 PRELIMINARY; PRT; 441 AA.  
AC Q8ZFP6  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Probable histidine acid phosphatase (EC 3.1.3.2) (Phosphoanhydride phosphorylase).  
GN YP01648 OR APPA OR Y1810.  
OS Yersinia pestis.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Yersinia.  
OX NCBI\_TaxID=632;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CO-92 / Biovar Orientalis;  
RX MEDLINE=21470413; PubMed=11586360;  
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,  
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,  
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,  
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,  
RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,  
RA Leather S., Moulé S., Oyston P.C.F., Quail M., Rutherford K.,  
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;  
RT "Genome sequence of Yersinia pestis, the causative agent of plague."  
RL Nature 413:523-527(2001).  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=KIM5 / Biovar Mediaevalis;  
RX MEDLINE=22137863; PubMed=12142430;  
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,  
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,  
RA Festerstein J.D., Lindler L.E., Brubaker R.W., Plano G.V.,  
RA Straley S.C., McDonough K.A., Nilles M.D., Matson J.S., Blattner F.R.,  
RA Perry R.D.;  
RT "Genome sequence of Yersinia pestis KIM."  
RL J. Bacteriol. 184:4601-4611(2002).  
DR EMBL; AJ414149; CAC90470.1; --  
DR EMBL; AF013783; AAM85378.1; --  
DR PIR; AC0201; AC0201.  
DR GO; GO:0003993; F:acid phosphatase activity; IEA.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR000560; HisAc\_phsphtase.  
DR Pfam; PF00328; acid\_phosphat\_1.  
DR PROSITE; PS00616; HIS\_ACID\_PHOSPHAT\_1; 1.  
DR Hydrolase; Complete proteome.  
KW SEQUENCE 441 AA; 47784 MW; CDD15COA64C98439 CRC64;

Query Match 44.3%; Score 954; DB 16; Length 441;  
Best Local Similarity 46.8%; Pred. No. 1.7e-71;  
Matches 195; Conservative 65; Mismatches 145; Indels 12; Gaps 6;  
QY 2 SEPE-LKLESVVVSRHGVRAP-TKATQLMQDVTPDAMPVTKVKGELTPGGELIAYLGH 60  
DB 29 AEPGTYLLEVRVILSRHGVRAP-TKATQLMQDVTPDAMPVTKVKGELTPGGELIAYLGH 88  
QY 61 YWRQRLVADGLLPKCGCPQSQGVVAIADVDERTKTEGFAAGLAPDCAITVHTQADTSS 120  
DB 89 FYGDFYRSLGLL-AAGCPAEGVVAQADIDQRTLTGQAFLDGVA PCGGLTVHQAQLKX 147  
QY 121 PDPLFNPLKTVGVCOLDNANVTDAILERAGGSIADFTGHYQTAFARELRLVNFPSNCLK 180  
DB 148 TDPLFHPVEAGVCKLDAQAOTDKAIEQLGGSLDVTVSQRYAKPPAQMGDVLNFAASPYCKS 207  
QY 181 REKQDESCLTAQALPSELKVSAD--CVSLTGAVALSLMLTEIFLLQAQGMPEPGWRIT 238  
DB 208 LQQQKTCDFAFHFAENVNKEGTVKVLSPALSLSTLGEIFLLQNAQMPVAKQRLX 267  
QY 239 DSHQNTLLSLHNAQFDLLQRTPEVARSRATPLDLIKNTALTPHPQKQAVGVLTPS-- 296  
DB 268 GAENWVSLSLHNAQFNLMAKTPVIARHKGTFLPQQIDTALT---LQIDAQGGKLPISAQ 324

QY 297 --VLEIAGHDTNLANLGALLENMTLPQGPONTPPGGELVFERWRLRLSDNSQWIOVSLVF 354  
DB 325 NEVFLGCHDTNLANIAGWLGADWQLPEQPDNTPPGGGLVFWLQNDHQRYVAKMFY 384  
QY 355 QTLQWRDKTPLSL-NTTPGVEVKLTLAGECBNAQMGCSLAGFTQIVNEARIPACSL 410  
DB 385 QTMQDLNRNAEKLDLKNPAGIISVAVAGCENNGDDKLCELDFTFQKVKVATIEPACHI 441  
RESULT 13  
Q8XBZ6 PRELIMINARY; PRT; 413 AA.  
AC Q8XBZ6  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Periplasmic glucose-1-phosphatase.  
GN AGP OR Z1421 OR ECS1158  
OS Escherichia coli O157:H7.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=83334;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
RX MEDLINE=21074935; PubMed=11206551;  
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blattner F.R.;  
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."  
RL Nature 409:529-533(2001).  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / RIMD 0509952;  
RX MEDLINE=21156231; PubMed=11258796;  
RA Hayaishi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tohe T.,  
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12."  
RL DNA Res. 8:11-22(2001).  
DR EMBL; AE005294; AAG55550.1; --  
DR EMBL; AP002554; BAB34581.1; --  
DR PIR; B85636; B85636.  
DR PIR; F90773; F90773.  
DR GO; GO:0003993; F:acid phosphatase activity; IEA.  
DR InterPro; IPR000560; HisAc\_phsphtase.  
DR Pfam; PF00328; acid\_phosphat\_1.  
DR PROSITE; PS00616; HIS\_ACID\_PHOSPHAT\_1; 1.  
DR PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.  
KW Complete proteome.  
QY SEQUENCE 413 AA; 45640 MW; A4630644EF6FE1A CRC64;

Query Match 25.6%; Score 551; DB 16; Length 413;  
Best Local Similarity 33.3%; Pred. No. 1e-37;  
Matches 138; Conservative 71; Mismatches 169; Indels 36; Gaps 12;  
QY 1 QSEPE-LKLESVVVSRHGVRAP-TKATQLMQDVTPDAMPVTKVKGELTPGGELIAYL 58  
DB 23 QTVPEGYQQVQLVMSRHLNLRAPLANNGSVLEQSTFNKWPWDVPGQJTTKGGVLEVYM 82  
QY 59 GHYWRQRLVADGLLPKCGCPQSQGVVAIADVDERTKTEGFAAGLAPDCAITVHTQADT 118  
DB 83 GHYWRQRLVADGLLPKCGCPQSQGVVAIADVDERTKTEGFAAGLAPDCAITVHTQADT 142  
QY 119 SPDPDLFNPLKTVGVCOLDNANVTDAILERAGGSIADFTGHYQTAFARELRLVNFPS 175  
DB 143 GTMDPTFNPVITD---DSAAFSEQAVAMEKSLK-QLTDSYQL---LEKIVNYKDS 193

QY 176 NLCLKREKQDESCSLTQALPSSELKVSDCVSLTGAVSLASMLTEIFLLQQAGMP--EPG 233  
DB :  
DQ 194 PAC-----KEKQCSSLVDCKNTFSAKYQCEPGVGLKVGNSLVDATFTLTQTYEGFPMDQVA 249  
DY :  
QY 234 WGRITDSHQWNTLLSHNAQFDLLORTPEVARSRATPLDLLIKTALTTPHPKPQAQGVTL 293  
DB :  
DQ 250 WGEIKSQDQWKVL SKLRNGYQDSLFTSPFARNVAKLPYSIDKAL-----VTD 298  
DY :  
QY 294 PTS----VLFIAGHTNLANLGALLLN-WTLFGQPDPNTPPGGELVFPERWRRLSDNSQM 349  
DB :  
DQ 299 RTSAPKITIVGHDSNIASLLTALDFRXYQLHDQNERTPIGGKTIVQRWRDSCANRDLMK 358  
DY :  
QY 350 VLSLVQTTLQMRDKTPTSLINTPGEVKLTLAGCERNAAQMCSLAFTQTVNEA 403  
DB :  
DQ 359 IEYVTQSABEQLENADALTLOPAQRVTLELSSGC-PIDADGFCDMXPDSVLNEA 411  
DY :

RESULT 14

Q8Z7P1 PRELIMINARY; PRT: 413 AA.

ID Q8Z7P1 ID AC Q8Z7P1;  
DT DT 01-MAR-2002 (TrEMBLrel\_20, Created)  
DD DD 01-MAR-2002 (TrEMBLrel\_20, Last sequence update)  
DE DE 01-JUN-2003 (TrEMBLrel\_24, Last annotation update)  
DN DN Glucose-1-phosphatase (GlPase), secreted.  
OS STYL153 OR AGP OR Tl803.  
OC Salmonella typhi.  
OX Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OY Enterobacteriaceae; Salmonella.  
NCBI TaxID=601,  
[1]  
SEQUENCE FROM N.A.  
RP STRAIN=CT18;  
RX MEDLINE=21534947; PubMed=11677608;  
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J., Churher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M., Baker S., Basham D., Brooks K., Chillingworth T., Connerton P., Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J., Feltwell T., Hamlin N., Hogue A., Hien T.T., Holtroyd S., Jagels K., Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C., Quail M.K., Rutherford K., Simmonds M., Skelton J., Stevens K., Whitehead S., Barrrell B.G.;  
RA "Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18.";  
RL Nature 413:848-852(2001).  
RN [2]

RN SEQUENCE FROM N.A.  
RP STRAIN=Ty2 / ATCC 700931;  
RX MEDLINE=22531367; PubMed=12644504;  
RA Deng W., Licu S.-R., Plunkett G. III, Mayhew G.F., Rose D.J., Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;  
RA "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2 and CT18";  
RL J. Bacteriol. 185:2330-2337(2003).  
DR ENBL; AL627269; CAD08242.1; -.  
DR EMBL; AR016840; AAO69425.1; -.  
DR GO; GO:0003993; F:acid phosphatase activity; IEA.  
DR Inter-Pro; IPR000560; HisAc Phsphtase.  
DR Pfam; PF00328; acid phosphat-.  
DR PROSITE; PS00616; HIS\_ACID\_PHOSPHTAT\_1; 1.  
DR PROSITE; PS00778; HIS\_ACID\_PHOSPHTAT\_2; 1.  
KW Complete proteome.  
SQ SEQUENCE 413 AA; 45630 MW; E9EF563CBBD9D26 CRC64;

Query Match 25.2%; Score 543; DB 16; Length 413;  
Best Local Similarity 33.6%; Pred No. 4.8e-37;  
Matches 137; Conservative 66; Mismatches 181; Indels 24; Gaps 10

Qy	59	GHYWRQRLVADGLPLKCGCPOSGOVALIADVDRTRKTGEAPAGIAPDCAITVHTQADT	111
Db	83	GHYTRWLVADGLPLKCGCPOSGOVALIADVDRTRKTGEAPAGIAPDCAITVHTQADT	142
Qy	119	SSPDPFLNPLKGTGVCOLDNANVTDAILERAGGSIADFTGHYATFARELERVLNFPQSNLC	178
Db	143	GTWDTFFNPVITDSSAARQAOVA-MEKARSQI-----HLDESAYKLLSQIITHYQDSFSC	196
Qy	179	LKREKQDSCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLQQAQGMPP-EPGWGR	236
Db	197	----KEKHQCSLIDAKDTFSANYQOEPGVGQPLKVGNSLVDAFTLQYEGFPMQVAMGG	252
Qy	237	ITDSHOWNTLLSHNAQFDLLQRTPEVARGRATPLLDLTKTALTTPPKQOQAYGVTLPST	296
Db	253	IHTDQWKVLSKKNQYDVSFTSPTVARNVAAPLVKYIDKLV-----AERVSAP-K	304
Qy	297	VLFIAGHDTNLANTGGALELN-WTLPGQPDNTPPGELVFERRRRLUSDNSQIVSLVFO	355
Db	305	VTVLVGHDSNTASLITLIDFKFYQHQDYERTIGGQVFORVHWDGNARDLMKIEYVYQ	364
Qy	356	TLQQRDKTFLSLNTPGCEVKTLIAGCEERNAQMGSLAGTTOIVNEA	403
Db	365	SARQURNAEALTKSPAQRVTLLEUKGPV-DANGFCLDKFDNVNMTA	411
RESULT 15			
Q7UD02	Q7UD02	PRELIMINARY;	PRT: 413 AA.
AC	Q7UD02;		
DT	01-OCT-2003 (TREMBLrel. 25, Created)		
DT	01-OCT-2003 (TREMBLrel. 25, Last sequence update)		
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)		
DE	Periplasmic glucose-1-phosphatase.		
GN	AGP OR sl075		
OS	Shigella flexneri.		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
OC	Enterobacteriaceae; Shigella.		
OX	NCBI_TaxID=623;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=2457T / ATCC 700930 / Serotype 2a;		
EX	MEDLINE=22590274; PubMed=12704152;		
RA	Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,		
RA	Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,		
RA	Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,		
RA	Schwartz D.C., Blattner F.R.;		
RT	Complete genome sequence and comparative genomics of Shigella		
RT	flexneri serotype 2a strain 2457T."		
RL	Infect. Immun. 71:2775-2786(2003).		
DR	EMBL; AB016981; AAP16516.1; --		
SQ	SEQUENCE 413 AA; 45648 MW; 6996CAE70CC65B39 CRC64;		
Query Match 25.2%; Score 543; DB 16; Length 413;			
Best Local Similarity 32.8%; Pred. No. 4.8e-37;			
Matches 135; Conservative 73; Mismatches 173; Indels 30; Gaps 12;			
Qy	1	QSEPE-LKLESVVIVSRHGVRAP-TKATQLMQDVTTPDAWTPVKLGELTPRGELIAYL	58
Db	23	QTVPEGYQLQVVLNMRNLRAPLANNVSVLEQSTPNKWPENVDVPGGLTKGGVLEVM	82
Qy	59	GHYWRQRLVADGLPLKCGCPOSGOVALIADVDRTRKTGEAPAGIAPDCAITVHTQADT	118
Db	83	GHYWRQRLVADGLPLKCGCPOSGOVALIADVDRTRKTGEAPAGIAPDCAITVHTQADT	142
Qy	119	SSPDPFLNPLKGTGVCOLDNANVTD---AILERAGGSIADFTGHYATFARELERVLNFPQSN	175
Db	143	GTWDTFFNPVITDSSAARQAOVA-MEKARSQI-----HLDESAYKLLSQIITHYQDSFSC	196
Qy	176	NLCLEKEKQDSCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLQQAQGMPP-EPG	233
Db	194	PAC-----KEKQCSLIDAKDTFSANYQOEPGVGQPLKVGNSLVDAFTLQYEGFPMQVAMGG	249
Qy	234	WGRITDSHOWNTLLSHNAQFDLLQRTPEVARGRATPLLDLTKTALTTPPKQOQAYGVTLP	293

```

Db      250  WGBIKSPQONKVLKXNGYQDSLFTSPEVARNVAKPLVSYIDKALV--TDRASAPKITV 307
Qy      294  PTVLFIAGHDTNLANIGGALELN-WLPGQPDNTPGGELVFERWRRLSDNSQWIOVSL 352
Db      308  -----LVGHSDNIALSLTALDFKPYQLHDQNEBTPIGGKIVFORWHDKXANRDLMKIEY 361
Qy      353  VFQLOQMRDTPISLNTPEGEVKLTLAGCEBERNAQCMCSLAGFTQIVNEA 403
Db      362  VYQSAEQLRNADALTLOAPARQVTLLESGC-PIDANGFCFMDKFDVSVLNEA 411

```

Search completed: April 30, 2004, 12:41:25  
 Job time : 38.1429 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 12:41:31 ; Search time 40 Seconds  
(without alignments)  
2841.193 Million cell updates/sec

Title: US-10-021-723B-13  
Perfect score: 2153  
Sequence: 1 QSEPELKLESVIVSRHGVR.....CSLAGTQIVNEARIPACSL 410

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1138120 seqs, 277189581 residues

Total number of hits satisfying chosen parameters: 1138120

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2153	100.0	410	14	US-10-021-723A-13
2	2153	100.0	410	14	US-10-021-723A-13
3	2153	100.0	412	14	US-10-334-672-1
4	2153	100.0	412	14	US-10-334-671-1
5	2153	100.0	436	14	US-10-156-660-2
6	2123	98.6	430	9	US-09-866-379-10
7	2077	96.5	432	9	US-09-866-379-8
8	2077	96.5	432	12	US-10-282-122A-43351
9	2077	96.5	432	14	US-10-156-660-4
10	2066	96.0	440	9	US-09-777-566A-2
11	2066	96.0	440	9	US-09-866-379-2
12	2066	96.0	440	14	US-10-034-985-2
13	2066	96.0	440	15	US-10-430-356-2
14	2042	94.8	432	15	US-10-284-962-3
15	2042	94.8	432	15	US-10-284-962-14

15	2042	94.8	433	14	US-10-266-041-1	Sequence 1, Appli
17	2039	94.7	432	15	US-10-284-962-5	Sequence 5, Appli
18	2007	93.2	432	14	US-10-021-723A-16	Sequence 15, Appli
19	971	45.1	441	14	US-10-021-723A-12	Sequence 12, Appli
20	954	44.3	441	12	US-10-282-122A-77792	Sequence 77792, A
21	946	43.9	441	14	US-10-021-723A-2	Sequence 2, Appli
22	943	43.6	420	14	US-10-021-723A-4	Sequence 4, Appli
23	773.5	35.9	476	14	US-10-021-723A-10	Sequence 10, Appli
24	613.5	28.5	318	14	US-10-021-723A-14	Sequence 14, Appli
25	463.5	21.5	409	14	US-10-021-723A-8	Sequence 8, Appli
26	458.5	21.3	421	14	US-10-021-723A-6	Sequence 6, Appli
27	184	8.5	426	12	US-10-257-174-44	Sequence 44, Appli
28	184	8.5	426	12	US-10-343-357-6	Sequence 6, Appli
29	122.5	5.7	99	12	US-10-282-122A-59099	Sequence 59099, A
30	119.5	5.6	467	12	US-10-442-538-165	Sequence 165, App
31	118.5	5.5	439	14	US-10-062-848-3	Sequence 3, Appli
32	118.5	5.5	440	12	US-10-442-538-106	Sequence 106, App
33	118.5	5.5	440	12	US-10-442-538-128	Sequence 128, App
34	118.5	5.5	440	12	US-10-442-538-151	Sequence 151, App
35	118.5	5.5	449	14	US-10-062-848-12	Sequence 12, Appli
36	118.5	5.5	465	13	US-10-083-452-8	Sequence 8, Appli
37	118.5	5.5	465	14	US-10-062-848-78	Sequence 78, Appli
38	118.5	5.5	465	14	US-10-229-358-6	Sequence 6, Appli
39	118.5	5.5	474	14	US-10-213-390-24	Sequence 24, Appli
40	117.5	5.5	440	12	US-10-442-538-107	Sequence 107, App
41	117.5	5.5	440	12	US-10-442-538-129	Sequence 129, App
42	117.5	5.5	440	12	US-10-442-538-152	Sequence 152, App
43	110.5	5.1	440	12	US-10-442-538-108	Sequence 108, App
44	110.5	5.1	440	12	US-10-442-538-130	Sequence 130, App
45	110.5	5.1	440	12	US-10-442-538-153	Sequence 153, App

## ALIGNMENTS

RESULT 1  
US-10-021-723A-13  
; Sequence 13, Application US/10021723A  
; Publication No. US20030101476A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay  
; APPLICANT: Mathur, Eric  
; APPLICANT: Richardson, Toby  
; APPLICANT: Robertson, Dan  
; APPLICANT: Barton, Nelson  
; TITLE OF INVENTION: Recombinant Phytases and Uses Thereof  
; FILE REFERENCE: 112766.140 (DIV-016CIP)  
; CURRENT APPLICATION NUMBER: US/10/021,723A  
; CURRENT FILING DATE: 2002-10-23  
; PRIOR APPLICATION NUMBER: US 60/255,090  
; PRIOR FILING DATE: 2000-12-12  
; NUMBER OF SEQ ID NOS: 74  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13  
; LENGTH: 410  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: amino acid sequence of a phytase enzyme with  
; OTHER INFORMATION: predicted glycosylation sites  
US-10-021-723A-13

Query Match	100.0%;	Score 2153;	DB 14;	Length 410;
Best Local Similarity	100.0%;	Pred. No. 2.3e-199;		
Matches 410;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	QSEPELKLESVIVSRHGVRAPTKATQMQDTPDANPTWPKLGLTPRGELIAYLGH	60	
Db	1	QSEPELKLESVIVSRHGVRAPTKATQMQDTPDANPTWPKLGLTPRGELIAYLGH	60	
Qy	61	YWRQRLVADGLPKCCPGSGQVAILADYDTRTKTGEAFAAGLAPDCAITVHTQADTSS	120	
Db	61	YWRQRLVADGLPKCCPGSGQVAILADYDTRTKTGEAFAAGLAPDCAITVHTQADTSS	120	

QY 121 PDPLFNPLKTVGCOLDNANVTDAILERAGGSIADFTGHYQTAFARELERVLPQSNLCLK 180  
DB 121 PDPLFNPLKTVGCOLDNANVTDAILERAGGSIADFTGHYQTAFARELERVLPQSNLCLK 180  
QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 240  
DB 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 240  
QY 241 HOWNTLLSLHNAQFDLLQRTPEVARSRATPLDLIKTALTALTPHPPOKQAYGVTLPSTVLFI 300  
DB 241 HOWNTLLSLHNAQFDLLQRTPEVARSRATPLDLIKTALTALTPHPPOKQAYGVTLPSTVLFI 300  
QY 301 AGHDNTNLNLGGALELAWNTLPGQDNTPPGGELVFERWRRLSDNSQWIOVSLVFTQLQOM 360  
DB 301 AGHDNTNLNLGGALELAWNTLPGQDNTPPGGELVFERWRRLSDNSQWIOVSLVFTQLQOM 360  
QY 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 410  
DB 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 410

## RESULT 2

US-10-021-723A-15

; Sequence 15, Application US/10021723A

; Publication No. US20030101476A1

; GENERAL INFORMATION:

; APPLICANT: Mathur, Jay

; APPLICANT: Richardson, Toby

; APPLICANT: Robertson, Dan

; APPLICANT: Barton, Nelson

; TITLE OF INVENTION: Recombinant Phytases and Uses Thereof

; FILE REFERENCE: 112766.140 (DIV-016CIP)

; CURRENT APPLICATION NUMBER: US/10/021,723A

; CURRENT FILING DATE: 2002-10-23

; PRIOR APPLICATION NUMBER: US 60/255,090

; PRIOR FILING DATE: 2000-12-12

; NUMBER OF SEQ ID NOS: 74

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 15

; LENGTH: 410

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: amino acid sequence of a phytase enzyme with

; OTHER INFORMATION: Predicted glycosylation sites

US-10-021-723A-15

Query Match 100.0%; Score 2153; DB 14; Length 410;  
Best Local Similarity 100.0%; Pred. No. 2.3e-199;  
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTDPDAMPTWPKLGELTPRGGELIAYLGH 60  
DB 1 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTDPDAMPTWPKLGELTPRGGELIAYLGH 60  
QY 61 YWRQLVADGLLPKCGCPQSGQVAILADVDERTRKTGEAFAGLAPDCAITVHTQADTSS 120  
DB 61 YWRQLVADGLLPKCGCPQSGQVAILADVDERTRKTGEAFAGLAPDCAITVHTQADTSS 120  
QY 121 PDPLFNPLKTVGCOLDNANVTDAILERAGGSIADFTGHYQTAFARELERVLPQSNLCLK 180  
DB 121 PDPLFNPLKTVGCOLDNANVTDAILERAGGSIADFTGHYQTAFARELERVLPQSNLCLK 180  
QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 240  
DB 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 240  
QY 241 HOWNTLLSLHNAQFDLLQRTPEVARSRATPLDLIKTALTALTPHPPOKQAYGVTLPSTVLFI 300  
DB 241 HOWNTLLSLHNAQFDLLQRTPEVARSRATPLDLIKTALTALTPHPPOKQAYGVTLPSTVLFI 300

QY 301 AGHDNTNLNLGGALELAWNTLPGQDNTPPGGELVFERWRRLSDNSQWIOVSLVFTQLQOM 360  
DB 301 AGHDNTNLNLGGALELAWNTLPGQDNTPPGGELVFERWRRLSDNSQWIOVSLVFTQLQOM 360  
QY 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 410  
DB 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 410

## RESULT 3

US-10-334-672-1

; Sequence 1, Application US/10334672

; Publication No. US20030157646A1

; GENERAL INFORMATION:

; APPLICANT: Lanahan, Mike

; APPLICANT: Koepf, Edward

; APPLICANT: Kretz, Keith

; TITLE OF INVENTION: Microbially-Expressed Thermotolerant Phytase For Animal Feed

; FILE REFERENCE: SYNG-P01-001

; CURRENT APPLICATION NUMBER: US/10/334,672

; CURRENT FILING DATE: 2002-12-30

; PRIOR APPLICATION NUMBER: 60/344,523

; PRIOR FILING DATE: 2001-12-28

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 1

; LENGTH: 412

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: No. US20030157646A19X Phytase

US-10-334-672-1

Query Match 100.0%; Score 2153; DB 14; Length 412;  
Best Local Similarity 100.0%; Pred. No. 2.3e-199;  
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTDPDAMPTWPKLGELTPRGGELIAYLGH 60  
DB 3 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTDPDAMPTWPKLGELTPRGGELIAYLGH 62  
QY 61 YWRQLVADGLLPKCGCPQSGQVAILADVDERTRKTGEAFAGLAPDCAITVHTQADTSS 120  
DB 63 YWRQLVADGLLPKCGCPQSGQVAILADVDERTRKTGEAFAGLAPDCAITVHTQADTSS 122  
QY 121 PDPLFNPLKTVGCOLDNANVTDAILERAGGSIADFTGHYQTAFARELERVLPQSNLCLK 180  
DB 123 PDPLFNPLKTVGCOLDNANVTDAILERAGGSIADFTGHYQTAFARELERVLPQSNLCLK 182  
QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 240  
DB 183 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 242  
QY 241 HOWNTLLSLHNAQFDLLQRTPEVARSRATPLDLIKTALTALTPHPPOKQAYGVTLPSTVLFI 300  
DB 243 HOWNTLLSLHNAQFDLLQRTPEVARSRATPLDLIKTALTALTPHPPOKQAYGVTLPSTVLFI 302  
QY 301 AGHDNTNLNLGGALELAWNTLPGQDNTPPGGELVFERWRRLSDNSQWIOVSLVFTQLQOM 360  
DB 303 AGHDNTNLNLGGALELAWNTLPGQDNTPPGGELVFERWRRLSDNSQWIOVSLVFTQLQOM 362  
QY 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 410  
DB 363 RDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 412

## RESULT 4

US-10-334-671-1

; Sequence 1, Application US/10334671

; Publication No. US20030170293A1

; GENERAL INFORMATION:

; APPLICANT: Lanahan, Mike

; APPLICANT: Koepf, Edward

APPLICANT: Kretz, Keith  
TITLE OF INVENTION: Microbially-Expressed Thermotolerant Phytase For Animal Feed  
FILE REFERENCE: 70098  
CURRENT APPLICATION NUMBER: US/10/334,671  
CURRENT FILING DATE: 2002-12-30  
PRIOR APPLICATION NUMBER: 60/344,523  
PRIOR FILING DATE: 2001-12-28  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: Patent in version 3.1  
SEQ ID NO 1  
LENGTH: 412  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: No. US20030170293A19X Phytase  
US-10-334-671-1

Query Match 100.0%; Score 2153; DB 14; Length 412;  
Best Local Similarity 100.0%; Pred. No. 2.3e-199;  
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QSEPELKLESVIVSRHGVRAPTKATQLMQDVTPDAMPTWPKLGELTPRGGLIAYLGH 60  
DB 3 QSEPELKLESVIVSRHGVRAPTKATQLMQDVTPDAMPTWPKLGELTPRGGLIAYLGH 62  
QY 61 YWRQRLVADGLLPKCGCPQSGQVAIIADVDERTKGTGEAFAAGLAPDCAITVHTQADTSS 120  
DB 63 YWRQRLVADGLLPKCGCPQSGQVAIIADVDERTKGTGEAFAAGLAPDCAITVHTQADTSS 122  
QY 121 PDPLFNPLKTVGCQLDNANVTDAILERAGGSADFTGHYQTAFARELERVLPQSNLCLK 180  
DB 123 PDPLFNPLKTVGCQLDNANVTDAILERAGGSADFTGHYQTAFARELERVLPQSNLCLK 182  
QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLQQAQGMPEPGWGRITDS 240  
DB 183 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLQQAQGMPEPGWGRITDS 242  
QY 241 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLDLIKTALTPHPPQKQAYGVTLPTSVLFI 300  
DB 243 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLDLIKTALTPHPPQKQAYGVTLPTSVLFI 302  
QY 301 AGHDTNLNLGAGLELNWTLPGQDNTPPGGELVFERWRRLSDNSQIQLSVFQTLQOM 360  
DB 303 AGHDTNLNLGAGLELNWTLPGQDNTPPGGELVFERWRRLSDNSQIQLSVFQTLQOM 362  
QY 361 RDKTPLSLNTPPEGVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 410  
DB 363 RDKTPLSLNTPPEGVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 412

RESULT 5  
US-10-156-660-2  
Sequence 2, Application US/10156660  
Publication No. US20030103958A1  
GENERAL INFORMATION:  
APPLICANT: Short, Jay M.  
APPLICANT: Kretz, Keith  
APPLICANT: Gray, Kevin A.  
APPLICANT: Barton, Nelson R.  
APPLICANT: Garrett, James B.  
APPLICANT: O'Donoghue, Eileen  
APPLICANT: Mathur, Eric J.  
TITLE OF INVENTION: PHYTASES, NUCLEIC ACIDS ENCODING THEM  
TITLE OF INVENTION: AND METHODS FOR MAKING AND USING THEM  
FILE REFERENCE: 09010-029007  
CURRENT APPLICATION NUMBER: US/10/156,660  
CURRENT FILING DATE: 2002-10-01  
PRIOR APPLICATION NUMBER: US 09/866,379  
PRIOR FILING DATE: 2001-05-24  
PRIOR APPLICATION NUMBER: US 09/580,515  
PRIOR FILING DATE: 2000-05-25  
PRIOR APPLICATION NUMBER: US 09/318,528  
PRIOR FILING DATE: 1999-05-25

PRIOR APPLICATION NUMBER: US 09/291,931  
PRIOR FILING DATE: 1999-04-13  
PRIOR APPLICATION NUMBER: US 09/259,214  
PRIOR FILING DATE: 1999-03-01  
PRIOR APPLICATION NUMBER: US 08/910,798  
PRIOR FILING DATE: 1997-08-13  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 436  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: modified phytase enzyme  
US-10-156-660-2

Query Match 100.0%; Score 2153; DB 14; Length 436;  
Best Local Similarity 100.0%; Pred. No. 2.6e-199;  
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QSEPELKLESVIVSRHGVRAPTKATQLMQDVTPDAMPTWPKLGELTPRGGLIAYLGH 60  
DB 23 QSEPELKLESVIVSRHGVRAPTKATQLMQDVTPDAMPTWPKLGELTPRGGLIAYLGH 82  
QY 61 YWRQRLVADGLLPKCGCPQSGQVAIIADVDERTKGTGEAFAAGLAPDCAITVHTQADTSS 120  
DB 83 YWRQRLVADGLLPKCGCPQSGQVAIIADVDERTKGTGEAFAAGLAPDCAITVHTQADTSS 142  
QY 121 PDPLFNPLKTVGCQLDNANVTDAILERAGGSADFTGHYQTAFARELERVLPQSNLCLK 180  
DB 143 PDPLFNPLKTVGCQLDNANVTDAILERAGGSADFTGHYQTAFARELERVLPQSNLCLK 202  
QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLQQAQGMPEPGWGRITDS 240  
DB 203 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLQQAQGMPEPGWGRITDS 262  
QY 241 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLDLIKTALTPHPPQKQAYGVTLPTSVLFI 300  
DB 263 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLDLIKTALTPHPPQKQAYGVTLPTSVLFI 322  
QY 301 AGHDTNLNLGAGLELNWTLPGQDNTPPGGELVFERWRRLSDNSQIQLSVFQTLQOM 360  
DB 323 AGHDTNLNLGAGLELNWTLPGQDNTPPGGELVFERWRRLSDNSQIQLSVFQTLQOM 382  
QY 361 RDKTPLSLNTPPEGVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 410  
DB 383 RDKTPLSLNTPPEGVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 432

RESULT 6  
US-09-866-379-10  
Sequence 10, Application US/09866379  
Patent No. US20020136754A1  
GENERAL INFORMATION:  
APPLICANT: DIVERSA CORPORATION  
APPLICANT: SHORT, Jay  
APPLICANT: KRETZ, Keith  
APPLICANT: GRAY, Kevin  
APPLICANT: BARTON, Nelson  
APPLICANT: GARRETT, James  
APPLICANT: O'DONOGHUE, Eileen  
TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYTASES AND USES THEREOF  
FILE REFERENCE: DIVER1370-7  
CURRENT APPLICATION NUMBER: US/09/866,379  
CURRENT FILING DATE: 2001-05-24  
PRIOR APPLICATION NUMBER: US 09/580,515  
PRIOR FILING DATE: 2000-05-25  
PRIOR APPLICATION NUMBER: US 09/318,528  
PRIOR FILING DATE: 1999-05-25  
PRIOR APPLICATION NUMBER: US 09/291,931  
PRIOR FILING DATE: 1999-04-13  
PRIOR APPLICATION NUMBER: US 09/259,214  
PRIOR FILING DATE: 1999-03-01

;; PRIOR APPLICATION NUMBER: US 08/910,798  
;; PRIOR FILING DATE: 1997-08-13  
;; NUMBER OF SEQ ID NOS: 10  
;; SOFTWARE: Patent in version 3.1  
;; SEQ ID NO 10  
;; LENGTH: 430  
;; TYPE: PRT  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: Modified phytase  
US-09-866-379-10

Query Match 98.6%; Score 2123; DB 9; Length 430;  
Best Local Similarity 99.5%; Pred. No. 2e-196;  
Matches 408; Conservative 0; Mismatches 0; Indels 2; Gaps 2;  
QY 1 QSEPELKLESVIVSRHGVRAPTKATQLMQDVTPDAMPTWPKVKGELTPRGELIAYLGH 60  
DB 23 QSEPELKLESVIVSRHGVRAPTKATQLMQDVTPDAMPTWPKVKGELTPRGELIAYLGH 82  
QY 61 YWRQLVADGLLPKCGCPQSGQVAIIADVDERTKRTGTEFAAGLAPDCAITVHTQADTSS 120  
DB 83 YWRQLVADGLLPKCGCPQSGQVAIIADVDERTKRTGTEFAAGLAPDCAITVHTQADTSS 142  
QY 121 PDPLFNPLKTGVCGOLDNANVTDAILERAGGSIAFTGHYQTAFARELERVLPQSNCLK 180  
DB 143 PDPLFNPLKTGVCGOLDNANVTDAILERAGGSIAFTGHYQTAFARELERVLPQSNCLK 201  
QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGMGRITDS 240  
DB 202 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGMGRITDS 261  
QY 241 HQWNTLLSHNAQFDLQRTPEVARSRAATPLLDLIKTALTTPHPKQKQAYGVTLPSTVLF 300  
DB 262 HQWNTLLSHNAQFDLQRTPEVARSRAATPLLDLIKTALTTPHPKQKQAYGVTLPSTVLF 320  
QY 301 AGHDTNLANLGALELNWTLPGQDNTPPGGELVFERWRRLSDNSQMIQVSLVFTLQOM 360  
DB 321 AGHDTNLANLGALELNWTLPGQDNTPPGGELVFERWRRLSDNSQMIQVSLVFTLQOM 380  
QY 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQCMCSLAGFTQIVNEARIPACSL 410  
DB 381 RDKTPLSLNTPPGEVKLTLAGCEERNAQCMCSLAGFTQIVNEARIPACSL 430

RESULT 7  
US-09-866-379-8  
;; Sequence 8, Application US/09866379  
;; Patent No. US20020136754A1  
;; GENERAL INFORMATION:  
;; APPLICANT: DIVERSA CORPORATION  
;; APPLICANT: SHORT, Jay  
;; APPLICANT: KRETZ, Keith  
;; APPLICANT: GRAY, Kevin  
;; APPLICANT: BARTON, Nelson  
;; APPLICANT: GARRETT, James  
;; APPLICANT: O'DONOGHUE, Eileen  
;; TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYTASES AND USES THEREOF  
;; FILE REFERENCE: DIVER1370-7  
;; CURRENT APPLICATION NUMBER: US/09/866,379  
;; CURRENT FILING DATE: 2001-05-24  
;; PRIOR APPLICATION NUMBER: US 09/580,515  
;; PRIOR FILING DATE: 2000-05-25  
;; PRIOR APPLICATION NUMBER: US 09/318,528  
;; PRIOR FILING DATE: 1999-05-25  
;; PRIOR APPLICATION NUMBER: US 09/291,931  
;; PRIOR FILING DATE: 1999-04-13  
;; PRIOR APPLICATION NUMBER: US 09/259,214  
;; PRIOR FILING DATE: 1999-03-01  
;; PRIOR APPLICATION NUMBER: US 08/910,798  
;; PRIOR FILING DATE: 1997-08-13  
;; NUMBER OF SEQ ID NOS: 10  
;; SOFTWARE: Patent in version 3.1

;; SEQ ID NO 8  
;; LENGTH: 432  
;; TYPE: PRT  
;; ORGANISM: Escherichia coli  
US-09-866-379-8  
Query Match 96.5%; Score 2077; DB 9; Length 432;  
Best Local Similarity 98.0%; Pred. No. 5.7e-192;  
Matches 402; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
QY 1 QSEPELKLESVIVSRHGVRAPTKATQLMQDVTPDAMPTWPKVKGELTPRGELIAYLGH 60  
DB 23 QSEPELKLESVIVSRHGVRAPTKATQLMQDVTPDAMPTWPKVKGELTPRGELIAYLGH 82  
QY 61 YWRQLVADGLLPKCGCPQSGQVAIIADVDERTKRTGTEFAAGLAPDCAITVHTQADTSS 120  
DB 83 YWRQLVADGLLPKCGCPQSGQVAIIADVDERTKRTGTEFAAGLAPDCAITVHTQADTSS 142  
QY 121 PDPLFNPLKTGVCGOLDNANVTDAILERAGGSIAFTGHYQTAFARELERVLPQSNCLK 180  
DB 143 PDPLFNPLKTGVCGOLDNANVTDAILERAGGSIAFTGHYQTAFARELERVLPQSNCLK 202  
QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGMGRITDS 240  
DB 203 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGMGRITDS 262  
QY 241 HQWNTLLSHNAQFDLQRTPEVARSRAATPLLDLIKTALTTPHPKQKQAYGVTLPSTVLF 300  
DB 263 HQWNTLLSHNAQFDLQRTPEVARSRAATPLLDLIKTALTTPHPKQKQAYGVTLPSTVLF 322  
QY 301 AGHDTNLANLGALELNWTLPGQDNTPPGGELVFERWRRLSDNSQMIQVSLVFTLQOM 360  
DB 323 AGHDTNLANLGALELNWTLPGQDNTPPGGELVFERWRRLSDNSQMIQVSLVFTLQOM 382  
QY 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQCMCSLAGFTQIVNEARIPACSL 410  
DB 383 RDKTPLSLNTPPGEVKLTLAGCEERNAQCMCSLAGFTQIVNEARIPACSL 432

RESULT 8  
US-10-282-122A-43351  
;; Sequence 43351, Application US/10282122A  
;; Publication No. US20040029129A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Wang, Liangsu  
;; APPLICANT: Zamudio, Carlos  
;; APPLICANT: Malone, Cheryl  
;; APPLICANT: Haselbeck, Robert  
;; APPLICANT: Onisen, Kari  
;; APPLICANT: Zyskind, Judith  
;; APPLICANT: Wall, Daniel  
;; APPLICANT: Trawick, John  
;; APPLICANT: Carr, Grant  
;; APPLICANT: Yamamoto, Robert  
;; APPLICANT: Forsyth, R.  
;; APPLICANT: Xu, H.  
;; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
;; FILE REFERENCE: ELITRA.034A  
;; CURRENT APPLICATION NUMBER: US/10/282,122A  
;; CURRENT FILING DATE: 2003-02-20  
;; PRIOR APPLICATION NUMBER: 60/191,078  
;; PRIOR FILING DATE: 2000-03-21  
;; PRIOR APPLICATION NUMBER: 60/206,848  
;; PRIOR FILING DATE: 2000-05-23  
;; PRIOR APPLICATION NUMBER: 60/207,727  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: 60/230,335  
;; PRIOR FILING DATE: 2000-09-06  
;; PRIOR APPLICATION NUMBER: 60/230,347  
;; PRIOR FILING DATE: 2000-09-09  
;; PRIOR APPLICATION NUMBER: 60/242,578  
;; PRIOR FILING DATE: 2000-10-23  
;; PRIOR APPLICATION NUMBER: 60/253,625



```
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 43351
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-282-122A-43351

Query Match          96.5%; Score 2077; DB 12; Length 432;
Best Local Similarity 98.0%; Pred. No. 5.7e-192;
Matches 402; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTPDANFTWPKLGELTPRGGLIAYLGH 60
DB 23 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTPDANFTWPKLGELTPRGGLIAYLGH 82
QY 61 YRQRLVADGLLPKCCPGSQGVAAIADVDERTRKTGEAFAAGLAPDCAITVHTQADTSS 120
DB 83 YRQRLVADGLLPKCCPGSQGVAAIADVDERTRKTGEAFAAGLAPDCAITVHTQADTSS 142
QY 121 PDPLENPLKTGVCOLDNANVTDAIILRAGGSADFTGHYQTAFRELERVLNFPQSNCLK 180
DB 143 PDPLENPLKTGVCOLDNANVTDAIILRAGGSADFTGHYQTAFRELERVLNFPQSNCLK 202
QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMPGPGWGRITDS 240
DB 203 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMPGPGWGRITDS 262
QY 241 HQWNTLLSHNAQFILLQRTPEVARSRATPLDLIKTALTTPHPPOKQAYGVTLPSTVLFI 300
DB 263 HQWNTLLSHNAQFILLQRTPEVARSRATPLDLIKTALTTPHPPOKQAYGVTLPSTVLFI 322
QY 301 AGHDTNLNLGALGNLWTLPGQDNTPPGGSLVFERWRRLSDNSQWIQVSLVFTQLQOM 360
DB 323 AGHDTNLNLGALGNLWTLPGQDNTPPGGSLVFERWRRLSDNSQWIQVSLVFTQLQOM 382
QY 361 RDKTPLSLNTPPGVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 410
DB 383 RDKTPLSLNTPPGVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 432
```

## RESULT 9

US-10-156-660-4

```
; Sequence 4, Application US/10156660
; Publication No. US20030103958A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay M.
; APPLICANT: Kretz, Keith
; APPLICANT: Gray, Kevin A.
; APPLICANT: Barton, Nelson R.
; APPLICANT: Garrett, James B.
; APPLICANT: O'Donoghue, Eileen
; APPLICANT: Mathur, Eric J.
; TITLE OF INVENTION: PHYTASES, NUCLEIC ACIDS ENCODING THEM
; TITLE OF INVENTION: AND METHODS FOR MAKING AND USING THEM
; FILE REFERENCE: 09010-029007
; CURRENT APPLICATION NUMBER: US/10156,660
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: US 09/866,379
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 09/580,515
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 09/291,931
```

```
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: US 09/259,214
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: US 08/910,798
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-156-660-4

Query Match          96.5%; Score 2077; DB 14; Length 432;
Best Local Similarity 98.0%; Pred. No. 5.7e-192;
Matches 402; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTPDANFTWPKLGELTPRGGLIAYLGH 60
DB 23 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTPDANFTWPKLGELTPRGGLIAYLGH 82
QY 61 YRQRLVADGLLPKCCPGSQGVAAIADVDERTRKTGEAFAAGLAPDCAITVHTQADTSS 120
DB 83 YRQRLVADGLLPKCCPGSQGVAAIADVDERTRKTGEAFAAGLAPDCAITVHTQADTSS 142
QY 121 PDPLENPLKTGVCOLDNANVTDAIILRAGGSADFTGHYQTAFRELERVLNFPQSNCLK 180
DB 143 PDPLENPLKTGVCOLDNANVTDAIILRAGGSADFTGHYQTAFRELERVLNFPQSNCLK 202
QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMPGPGWGRITDS 240
DB 203 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMPGPGWGRITDS 262
QY 241 HQWNTLLSHNAQFILLQRTPEVARSRATPLDLIKTALTTPHPPOKQAYGVTLPSTVLFI 300
DB 263 HQWNTLLSHNAQFILLQRTPEVARSRATPLDLIKTALTTPHPPOKQAYGVTLPSTVLFI 322
QY 301 AGHDTNLNLGALGNLWTLPGQDNTPPGGSLVFERWRRLSDNSQWIQVSLVFTQLQOM 360
DB 323 AGHDTNLNLGALGNLWTLPGQDNTPPGGSLVFERWRRLSDNSQWIQVSLVFTQLQOM 382
QY 361 RDKTPLSLNTPPGVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 410
DB 383 RDKTPLSLNTPPGVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 432
```

## RESULT 10

US-09-777-566A-2

```
; Sequence 2, Application US/09777566A
; Patent No. US20010055788A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: Short, Jay
; APPLICANT: Kretz, Keith
; TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYTASES AND USES THEREOF
; FILE REFERENCE: DIVER1370-6
; CURRENT APPLICATION NUMBER: US/09/777,566A
; CURRENT FILING DATE: 2001-06-11
; PRIOR APPLICATION NUMBER: US 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: US 09/259,214
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: US 08/910,798
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
```

NAME/KEY: misc.feature  
LOCATION: (1)..(1323)  
OTHER INFORMATION: n is any nucleotide  
US-09-777-566A-2

Query Match 96.0%; Score 2066; DB 9; Length 440;  
Best Local Similarity 97.6%; Pred. No. 6.8e-191;  
Matches 400; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
QY 1 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTPDAMPTWPKLGELTPRGGELIAYLGH 60  
DB 23 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTPDAMPTWPKLGELTPRGGELIAYLGH 82  
QY 61 YRQRLVADGLLPKCGCQSGQVAILIADVDERTRKTGEAFAAGLAPDCAITVHTQADTSS 120  
DB 83 YRQRLVADGLLPKCGCQSGQVAILIADVDERTRKTGEAFAAGLAPDCAITVHTQADTSS 142  
QY 121 PDPLFNPLKTGVCQLDNANVTDAILRAGGSIADFTGHQTAFARELERVLPQSNLCLK 180  
DB 143 PDPLFNPLKTGVCQLDNANVTDAILRAGGSIADFTGHQTAFARELERVLPQSNLCLK 202  
QY 181 REKQDESSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMPGPGHGRITDS 240  
DB 203 REKQDESSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMPGPGHGRITDS 262  
QY 241 HQWNTLLSHNAQFLLQRTPEVARSRATPLLDLTKALTTPHPPQKQAYGVTLPTSVLFI 300  
DB 263 HQWNTLLSHNAQFLLQRTPEVARSRATPLLDLTKALTTPHPPQKQAYGVTLPTSVLFI 322  
QY 301 AGHDTNLNLGALBELNWTLPQPDNTPPGGELVFERWRRLSDNSQIQLSVLFTQLOOM 360  
DB 323 AGHDTNLNLGALBELNWTLPQPDNTPPGGELVFERWRRLSDNSQIQLSVLFTQLOOM 382  
QY 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQCMCSLAGFTQIVNEARIPACSL 410  
DB 383 RDKTPLSLNTPPGEVKLTLAGCEERNAQCMCSLAGFTQIVNEARIPACSL 432

RESULT 11  
US-09-866-379-2  
Sequence 2, Application US/09866379  
Patent No. US20020136754A1  
GENERAL INFORMATION:  
APPLICANT: DIVERSA CORPORATION  
APPLICANT: SHORT, Jay  
APPLICANT: KRETZ, Keith  
APPLICANT: GRAY, Kevin  
APPLICANT: BARTON, Nelson  
APPLICANT: GARRETT, James  
APPLICANT: O'DONOGHUE, Eileen  
TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYTASES AND USES THEREOF  
FILE REFERENCE: DIVER1370-7  
CURRENT APPLICATION NUMBER: US/09/866,379  
PRIOR FILING DATE: 2001-05-24  
PRIOR APPLICATION NUMBER: US 09/580,515  
PRIOR FILING DATE: 2000-05-25  
PRIOR APPLICATION NUMBER: US 09/318,528  
PRIOR FILING DATE: 1999-05-25  
PRIOR APPLICATION NUMBER: US 09/291,931  
PRIOR FILING DATE: 1999-04-13  
PRIOR APPLICATION NUMBER: US 09/259,214  
PRIOR FILING DATE: 1999-03-01  
PRIOR APPLICATION NUMBER: US 08/910,798  
PRIOR FILING DATE: 1997-08-13  
NUMBER OF SEQ ID NOS: 10  
SOFTWARE: Patent in version 3.1  
SEQ ID NO 2  
LENGTH: 440  
TYPE: PRT  
ORGANISM: Escherichia coli  
US-09-866-379-2

Query Match 96.0%; Score 2066; DB 9; Length 440;

Best Local Similarity 97.6%; Pred. No. 6.8e-191;  
Matches 400; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
QY 1 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTPDAMPTWPKLGELTPRGGELIAYLGH 60  
DB 23 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTPDAMPTWPKLGELTPRGGELIAYLGH 82  
QY 61 YRQRLVADGLLPKCGCQSGQVAILIADVDERTRKTGEAFAAGLAPDCAITVHTQADTSS 120  
DB 83 YRQRLVADGLLPKCGCQSGQVAILIADVDERTRKTGEAFAAGLAPDCAITVHTQADTSS 142  
QY 121 PDPLFNPLKTGVCQLDNANVTDAILRAGGSIADFTGHQTAFARELERVLPQSNLCLK 180  
DB 143 PDPLFNPLKTGVCQLDNANVTDAILRAGGSIADFTGHQTAFARELERVLPQSNLCLK 202  
QY 181 REKQDESSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMPGPGHGRITDS 240  
DB 203 REKQDESSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMPGPGHGRITDS 262  
QY 241 HQWNTLLSHNAQFLLQRTPEVARSRATPLLDLTKALTTPHPPQKQAYGVTLPTSVLFI 300  
DB 263 HQWNTLLSHNAQFLLQRTPEVARSRATPLLDLTKALTTPHPPQKQAYGVTLPTSVLFI 322  
QY 301 AGHDTNLNLGALBELNWTLPQPDNTPPGGELVFERWRRLSDNSQIQLSVLFTQLOOM 360  
DB 323 AGHDTNLNLGALBELNWTLPQPDNTPPGGELVFERWRRLSDNSQIQLSVLFTQLOOM 382  
QY 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQCMCSLAGFTQIVNEARIPACSL 410  
DB 383 RDKTPLSLNTPPGEVKLTLAGCEERNAQCMCSLAGFTQIVNEARIPACSL 432

RESULT 12  
US-10-034-985-2  
Sequence 2, Application US/10034985  
Publication No. US20030049815A1  
GENERAL INFORMATION:  
APPLICANT: Kretz, Keith  
TITLE OF INVENTION: NOVEL PHYTASE  
FILE REFERENCE: 09010/029003  
CURRENT APPLICATION NUMBER: US/10/034,985  
CURRENT FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: US/09/580,515  
PRIOR FILING DATE: 1999-05-25  
PRIOR APPLICATION NUMBER: 09/291,931  
PRIOR FILING DATE: 1999-04-13  
PRIOR APPLICATION NUMBER: 08/910,798  
PRIOR FILING DATE: 1997-08-13  
PRIOR APPLICATION NUMBER: 09/259,214  
PRIOR FILING DATE: 1999-03-01  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 440  
TYPE: PRT  
ORGANISM: Escherichia coli  
US-10-034-985-2

Query Match 96.0%; Score 2066; DB 14; Length 440;  
Best Local Similarity 97.6%; Pred. No. 6.8e-191;  
Matches 400; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
QY 1 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTPDAMPTWPKLGELTPRGGELIAYLGH 60  
DB 23 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTPDAMPTWPKLGELTPRGGELIAYLGH 82  
QY 61 YRQRLVADGLLPKCGCQSGQVAILIADVDERTRKTGEAFAAGLAPDCAITVHTQADTSS 120  
DB 83 YRQRLVADGLLPKCGCQSGQVAILIADVDERTRKTGEAFAAGLAPDCAITVHTQADTSS 142  
QY 121 PDPLFNPLKTGVCQLDNANVTDAILRAGGSIADFTGHQTAFARELERVLPQSNLCLK 180  
DB 143 PDPLFNPLKTGVCQLDNANVTDAILRAGGSIADFTGHQTAFARELERVLPQSNLCLK 202

```
QY 181 REKQDECSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMPGPGWGRITDS 240
DB 203 REKQDECSLTQALPSELKVSADNVSLTGAVSLASMLTEIFLLQQAQMPGPGWGRITDS 262
QY 241 HQWNTLLSLHNAQFOLLQRTPEVARSRATPLDLIKTALTTPHPKQKQAYGVTLPTSVLFI 300
DB 263 HQWNTLLSLHNAQFYLLQRTPEVARSRATPLDLIIMALTTPHPKQKQAYGVTLPTSVLFI 322
QY 301 AGHDTNLANLGGALELNWTLPGQPDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 360
DB 323 AGHDTNLANLGGALELNWTLPGQPDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 382
QY 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 410
DB 383 RDKTPLSLNTPPGEVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 432

RESULT 13
US-10-430-356-2
; Sequence 2, Application US/10430356
; Publication No. US20030232041A1
; GENERAL INFORMATION:
; APPLICANT: Kretz, Keith
; TITLE OF INVENTION: NOVEL PHYTASE
; FILE REFERENCE: 09010/029003
; CURRENT APPLICATION NUMBER: US/10/430,356
; CURRENT FILING DATE: 2003-05-05
; PRIOR APPLICATION NUMBER: US/09/580,515
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: 08/910,798
; PRIOR FILING DATE: 1997-08-13
; PRIOR APPLICATION NUMBER: 09/259,214
; PRIOR FILING DATE: 1999-03-01
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-430-356-2

Query Match 96.0%; Score 2066; DB 15; Length 440;
Best Local Similarity 97.6%; Pred. No. 6.8e-191;
Matches 400; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 OSEPELKLESVVIVSRHGVRAPTATQALMODVTPDAWPTWPKLGTTPRGGLIAYLGH 60
DB 23 OSEPELKLESVVIVSRHGVRAPTATQALMODVTPDAWPTWPKLGTTPRGGLIAYLGH 82
QY 61 YWRQLVADGLLPKCGCPQSGQVAILIADVDERTKRTGEAFAAGLAPDCAITVHTQADTSS 120
DB 83 YWRQLVADGLLAKKCGCPQSGQVAILIADVDERTKRTGEAFAAGLAPDCAITVHTQADTSS 142
QY 121 PDPLFNPLKTGVCQLDNANVTDAILLRAGGSIADFTGHQTAFARELERVLPFQSNCLCK 180
DB 143 PDPLFNPLKTGVCQLDNANVTDAILLRAGGSIADFTGHQTAFARELERVLPFQSNCLCK 202
QY 181 REKQDECSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMPGPGWGRITDS 240
DB 203 REKQDECSLTQALPSELKVSADNVSLTGAVSLASMLTEIFLLQQAQMPGPGWGRITDS 262
QY 241 HQWNTLLSLHNAQFOLLQRTPEVARSRATPLDLIKTALTTPHPKQKQAYGVTLPTSVLFI 300
DB 263 HQWNTLLSLHNAQFYLLQRTPEVARSRATPLDLIIMALTTPHPKQKQAYGVTLPTSVLFI 322
QY 301 AGHDTNLANLGGALELNWTLPGQPDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 360
DB 323 AGHDTNLANLGGALELNWTLPGQPDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 382

RESULT 14
US-10-284-962-3
; Sequence 3, Application US/10284962
; Publication No. US20030206913A1
; GENERAL INFORMATION:
; APPLICANT: Weibel, Douglas M.
; APPLICANT: Orr, Donald E.
; APPLICANT: Ruch, Frank E.
; TITLE OF INVENTION: PHYTASE-CONTAINING ANIMAL FOOD AND METHOD
; FILE REFERENCE: 834460-71725
; CURRENT APPLICATION NUMBER: US/10/284,962
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: US 60/335,303
; PRIOR FILING DATE: 2001-10-31
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-284-962-3

Query Match 94.8%; Score 2042; DB 15; Length 432;
Best Local Similarity 96.6%; Pred. No. 1.4e-188;
Matches 396; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 QSEPELKLESVVIVSRHGVRAPTATQALMODVTPDAWPTWPKLGTTPRGGLIAYLGH 60
DB 23 QSEPELKLESVVIVSRHGVRAPTATQALMODVTPDAWPTWPKLGTTPRGGLIAYLGH 82
QY 61 YWRQLVADGLLPKCGCPQSGQVAILIADVDERTKRTGEAFAAGLAPDCAITVHTQADTSS 120
DB 83 YWRQLVADGLLAKKCGCPQSGQVAILIADVDERTKRTGEAFAAGLAPDCAITVHTQADTSS 142
QY 121 PDPLFNPLKTGVCQLDNANVTDAILLRAGGSIADFTGHQTAFARELERVLPFQSNCLCK 180
DB 143 PDPLFNPLKTGVCQLDNANVTDAILLRAGGSIADFTGHQTAFARELERVLPFQSNCLCK 202
QY 181 REKQDECSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMPGPGWGRITDS 240
DB 203 REKQDECSLTQALPSELKVSADNVSLTGAVSLASMLTEIFLLQQAQMPGPGWGRITDS 262
QY 241 HQWNTLLSLHNAQFOLLQRTPEVARSRATPLDLIKTALTTPHPKQKQAYGVTLPTSVLFI 300
DB 263 HQWNTLLSLHNAQFYLLQRTPEVARSRATPLDLIIMALTTPHPKQKQAYGVTLPTSVLFI 322
QY 301 AGHDTNLANLGGALELNWTLPGQPDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 360
DB 323 AGHDTNLANLGGALELNWTLPGQPDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 382
QY 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 410
DB 383 RDKTPLSLNTPPGEVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 432

RESULT 15
US-10-284-962-14
; Sequence 14, Application US/10284962
; Publication No. US20030206913A1
; GENERAL INFORMATION:
; APPLICANT: Weibel, Douglas M.
; APPLICANT: Orr, Donald E.
; APPLICANT: Ruch, Frank E.
; TITLE OF INVENTION: PHYTASE-CONTAINING ANIMAL FOOD AND METHOD
; FILE REFERENCE: 834460-71725
; CURRENT APPLICATION NUMBER: US/10/284,962
; CURRENT FILING DATE: 2002-10-31
```

```
QY 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 410
DB 383 RDKTPLSLNTPPGEVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 432

RESULT 14
US-10-284-962-3
; Sequence 3, Application US/10284962
; Publication No. US20030206913A1
; GENERAL INFORMATION:
; APPLICANT: Weibel, Douglas M.
; APPLICANT: Orr, Donald E.
; APPLICANT: Ruch, Frank E.
; TITLE OF INVENTION: PHYTASE-CONTAINING ANIMAL FOOD AND METHOD
; FILE REFERENCE: 834460-71725
; CURRENT APPLICATION NUMBER: US/10/284,962
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: US 60/335,303
; PRIOR FILING DATE: 2001-10-31
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-284-962-3

Query Match 94.8%; Score 2042; DB 15; Length 432;
Best Local Similarity 96.6%; Pred. No. 1.4e-188;
Matches 396; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 QSEPELKLESVVIVSRHGVRAPTATQALMODVTPDAWPTWPKLGTTPRGGLIAYLGH 60
DB 23 QSEPELKLESVVIVSRHGVRAPTATQALMODVTPDAWPTWPKLGTTPRGGLIAYLGH 82
QY 61 YWRQLVADGLLPKCGCPQSGQVAILIADVDERTKRTGEAFAAGLAPDCAITVHTQADTSS 120
DB 83 YWRQLVADGLLAKKCGCPQSGQVAILIADVDERTKRTGEAFAAGLAPDCAITVHTQADTSS 142
QY 121 PDPLFNPLKTGVCQLDNANVTDAILLRAGGSIADFTGHQTAFARELERVLPFQSNCLCK 180
DB 143 PDPLFNPLKTGVCQLDNANVTDAILLRAGGSIADFTGHQTAFARELERVLPFQSNCLCK 202
QY 181 REKQDECSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMPGPGWGRITDS 240
DB 203 REKQDECSLTQALPSELKVSADNVSLTGAVSLASMLTEIFLLQQAQMPGPGWGRITDS 262
QY 241 HQWNTLLSLHNAQFOLLQRTPEVARSRATPLDLIKTALTTPHPKQKQAYGVTLPTSVLFI 300
DB 263 HQWNTLLSLHNAQFYLLQRTPEVARSRATPLDLIIMALTTPHPKQKQAYGVTLPTSVLFI 322
QY 301 AGHDTNLANLGGALELNWTLPGQPDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 360
DB 323 AGHDTNLANLGGALELNWTLPGQPDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 382
QY 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 410
DB 383 RDKTPLSLNTPPGEVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 432

RESULT 15
US-10-284-962-14
; Sequence 14, Application US/10284962
; Publication No. US20030206913A1
; GENERAL INFORMATION:
; APPLICANT: Weibel, Douglas M.
; APPLICANT: Orr, Donald E.
; APPLICANT: Ruch, Frank E.
; TITLE OF INVENTION: PHYTASE-CONTAINING ANIMAL FOOD AND METHOD
; FILE REFERENCE: 834460-71725
; CURRENT APPLICATION NUMBER: US/10/284,962
; CURRENT FILING DATE: 2002-10-31
```

; PRIOR APPLICATION NUMBER: US 60/335,303  
 ; PRIOR FILING DATE: 2001-10-31  
 ; NUMBER OF SEQ ID NOS: 14  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 14  
 ; LENGTH: 432  
 ; TYPE: PRT  
 ; ORGANISM: Escherichia coli  
 US-10-284-962-14

Query Match	94.8%	Score 2042	DB 15	Length 432
Best Local Similarity	96.6%	Pred. No. 1.4e-188		
Matches 396	Conservative 0	Mismatches 14	Indels 0	Gaps 0

  

QY	1	QSEPELKLESVVIVSRHGVRAPTKATQIMQDVTTPDAMTPWVKIGELTPRGELIAYLGH	60
Db	23	QSEPELKLESVVIVSRHGVRAPTKATQIMQDVTTPDAMTPWVKIGELTPRGELIAYLGH	82
QY	61	YWRQLVADGLLPKCGCPQSGQVAVIADVDERTKRTGEAFAGLAPDCAITVHTQADTSS	120
Db	83	YWRQLVADGLLPKCGCPQSGQVAVIADVDERTKRTGEAFAGLAPDCAITVHTQADTSS	142
QY	121	PDPLFNPLKTGVCOLDNANVTDAIILERAGGSADFTGHYQTAFRELERVLAFFPQSNCLK	180
Db	143	PDPLFNPLKTGVCOLDNANVTDAIILSRAGGSADFTGHYQTAFRELERVLAFFPQSNCLK	202
QY	181	REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEFGMGRIIDS	240
Db	203	REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEFGMGRIIDS	262
QY	241	HOWNTLLSLHNAQFDLLQRTPEVARSRATPLLDLTKTALTTPHPPOKQAYGVTLPSTVLFI	300
Db	263	HOWNTLLSLHNAQFDLLQRTPEVARSRATPLLDLTKTALTTPHPPOKQAYGVTLPSTVLFI	322
QY	301	AGHDTNLANLGGALNLNWTLPQDPNTPPGSELVFERWRRLSDNSQWIQVSLVFQTLQOM	360
Db	323	AGHDTNLANLGGALNLNWTLPQDPNTPPGSELVFERWRRLSDNSQWIQVSLVFQTLQOM	382
QY	361	RDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL	410
Db	383	RDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL	432

Search completed: April 30, 2004, 12:55:03  
 Job time : 41 secs

GenCore version 5.1.6  
Copyright (C) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 12:36:16 ; Search time 16.4286 Seconds  
(without alignments)  
1288.405 Million cell updates/sec

Title: US-10-021-723B-13

Perfect score: 2153

Sequence: 1 QSEPELKLESVIVSRHGVR.....CSLAGTQIVNEARIPACSL 410

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*

- 1: /cgn2\_6/prodata/2/aaa/5A\_COMB.pep.\*
- 2: /cgn2\_6/prodata/2/aaa/5B\_COMB.pep.\*
- 3: /cgn2\_6/prodata/2/aaa/5A\_COMB.pep.\*
- 4: /cgn2\_6/prodata/2/aaa/6B\_COMB.pep.\*
- 5: /cgn2\_6/prodata/2/aaa/PCITUS\_COMB.pep.\*
- 6: /cgn2\_6/prodata/2/aaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2066	96.0	440	3	US-09-259-214-2
2	2066	96.0	440	3	US-09-318-528-2
3	2066	96.0	440	3	US-09-291-931-2
4	2042	94.8	433	4	US-09-540-149A-1
5	1958.5	91.0	423	2	US-08-910-798-2
6	539	25.0	421	4	US-09-489-039A-7512
7	370	17.2	522	4	US-09-489-039A-13501
8	119.5	5.6	467	4	US-09-684-855-165
9	118.5	5.5	439	4	US-09-044-718-3
10	118.5	5.5	440	4	US-09-684-855-106
11	118.5	5.5	440	4	US-09-684-855-128
12	118.5	5.5	440	4	US-09-684-855-151
13	118.5	5.5	449	4	US-09-044-718-12
14	118.5	5.5	465	3	US-08-868-435-33
15	118.5	5.5	465	4	US-08-744-231-33
16	118.5	5.5	465	4	US-09-044-718-78
17	118.5	5.5	465	4	US-09-636-499-6
18	117.5	5.5	465	4	US-09-273-871A-8
19	117.5	5.5	440	4	US-09-684-855-107
20	117.5	5.5	440	4	US-09-684-855-129
21	117.5	5.5	440	4	US-09-684-855-152
22	110.5	5.1	440	4	US-09-684-855-108
23	110.5	5.1	440	4	US-09-684-855-130
24	110.5	5.1	440	4	US-09-684-855-153
25	110.5	5.1	465	4	US-09-044-718-80
26	109.5	5.1	465	4	US-09-636-498-13
27	109.5	5.1	465	4	US-09-044-718-79

28	109.5	5.1	479	1	US-07-923-724-2	Sequence 2, Appli
29	109.5	5.1	479	2	US-08-609-426A-2	Sequence 2, Appli
30	109.5	5.1	479	2	US-08-374-522C-4	Sequence 4, Appli
31	108.5	5.0	440	4	US-09-684-855-109	Sequence 109, App
32	108.5	5.0	440	4	US-09-684-855-131	Sequence 131, App
33	108.5	5.0	440	4	US-09-684-855-154	Sequence 154, App
34	108.5	5.0	455	4	US-09-636-499-14	Sequence 14, Appl
35	108.5	5.0	465	4	US-09-044-718-81	Sequence 81, Appl
36	105	4.9	413	4	US-09-684-855-142	Sequence 142, App
37	105	4.9	422	4	US-09-684-855-119	Sequence 119, App
38	105	4.9	442	3	US-08-993-359-28	Sequence 28, Appl
39	105	4.9	442	4	US-09-273-871A-4	Sequence 4, Appli
40	105	4.9	442	4	US-09-482-558A-28	Sequence 28, Appl
41	103	4.8	1048	4	US-08-887-534A-85	Sequence 85, Appl
42	103	4.8	1048	4	US-09-527-431-85	Sequence 85, Appl
43	101.5	4.7	386	1	US-08-758-213-1	Sequence 1, Appli
44	101.5	4.7	386	2	US-08-692-787-48	Sequence 48, Appl
45	101.5	4.7	386	3	US-09-097-199-48	Sequence 48, Appl

#### ALIGNMENTS

RESULT 1  
US-09-259-214-2  
; Sequence 2, Application US/09259214A  
; Patent No. 6110719  
; GENERAL INFORMATION:  
; APPLICANT: Kretz, Keith  
; TITLE OF INVENTION: NOVEL PHRYASE  
; FILE REFERENCE: DIVER1370-1  
; CURRENT APPLICATION NUMBER: US/09/259,214A  
; CURRENT FILING DATE: 1999-03-01  
; EARLIER APPLICATION NUMBER: 08/910,798  
; EARLIER FILING DATE: 1997-08-13  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 440  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-259-214-2

Query Match	96.0%	Score	2066;	DB 3;	Length	440;			
Best Local Similarity	97.6%;	Pred. No.	2.3e-211;	Mismatches	10;	Indels	0;	Gaps	0;
Matches	400;	Conservative	0;						
Qy	1	QSEPELKLESVIVSRHGVRAPTKATQLMDVTPDAPWTPVKLGELTPRGGELIAYLGH	60						
Db	23	QSEPELKLESVIVSRHGVRAPTKATQLMDVTPDAPWTPVKLGELTPRGGELIAYLGH	82						
Qy	61	YKQRLVADGLPKCGCPQSGQVAIADVDERTKGTGEAPAGIADPADCAITVHTQADTSS	120						
Db	83	YQKRLVADGLLAKKCPQSGQVAIADVDERTKGTGEAPAGIADPADCAITVHTQADTSS	142						
Qy	121	PDPLNFKLTKGVQCLDNANVTDAILERAGGSIADFTGHVQTAFRELERVNFPOSNLCLK	180						
Db	143	PDPLNFKLTKGVQCLDNANVTDAILSRAGGSIADFTGHQRTAFRELERVNFPOSNLCLK	202						
Qy	181	REKQDSCSLTQALPSELKVSADCVSLTGVASLASMLTEIFLQOQGMPEPGWGRITDS	240						
Db	203	REKQDSCSLTQALPSELKVSADNVSLTGVASLASMLTEIFLQOQGMPEPGWGRITDS	262						
Qy	241	HWNTLLSLNNAQFDLLQRTPEVARSRATPLDLIKTALTPHPPOKQAYGVTLPTSVLFI	300						
Db	263	HWNTLLSLNNAQFYLLQRTPEVARSRATPLDLIMAAITPHPPQKQAYGVTLPTSVLFI	322						
Qy	301	AGHDTNLNIGGALELNWTLPGQPDNTPPGGELVFERWRRLSDNSQWISLVFQTLQOM	360						
Db	323	AGHDTNLNIGGALELNWTLPGQPDNTPPGGELVFERWRRLSDNSQWISLVFQTLQOM	382						
Qy	361	RDKTPLSLNTPPGEVKLTLAGCERNAQMGCSLAGFTQIVNEARIPACSL	410						

Db 383 RDKTPLSLNTPPGVRLTLGACERNAQGMCSLAGFTQIVNEARIPACSL 432

RESULT 2

US-09-318-528-2  
; Sequence 2, Application US/09318528  
; Patent No. 6183740  
; GENERAL INFORMATION:  
; APPLICANT: Kretz, Keith  
; TITLE OF INVENTION: NOVEL PHYTASE  
; FILE REFERENCE: 09010/029003  
; CURRENT APPLICATION NUMBER: US/09/318,528  
; CURRENT FILING DATE: 1999-05-25  
; EARLIER APPLICATION NUMBER: 09/291,931  
; EARLIER FILING DATE: 1999-04-13  
; EARLIER APPLICATION NUMBER: 08/910,798  
; EARLIER FILING DATE: 1997-08-13  
; EARLIER APPLICATION NUMBER: 09/259,214  
; EARLIER FILING DATE: 1999-03-01  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 440  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-318-528-2

Query Match 96.0%; Score 2066; DB 3; Length 440;  
Best Local Similarity 97.6%; Pred. No. 2.3e-211;  
Matches 400; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 1 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTTPDAPWTPVKLGELTPRGELIAYLGH 60  
Db 23 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTTPDAPWTPVKLGELTPRGELIAYLGH 82  
Qy 61 YMRQRLVADGLLPKCGCPOSGOVAIADVDETRTKTGEAFAAGLAPDCAITVHTQADTSS 120  
Db 83 YQRLVADGLLAKKCGCPOSGOVAIADVDETRTKTGEAFAAGLAPDCAITVHTQADTSS 142  
Qy 121 PDPLFNPLKTVGCOLDNANVTDAILLERAGSGIADFTGHYQTAFARELERVLPQSNCLCK 180  
Db 143 PDPLFNPLKTVGCOLDNANVTDAILLERAGSGIADFTGHYQTAFARELERVLPQSNCLCK 202  
Qy 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 240  
Db 203 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 262  
Qy 241 HOWNTLISLHNAQFDLLORTPEVARSRATPLDLIKTALTTPHPPOKQAYGVTLPSTVLF 300  
Db 263 HOWNTLISLHNAQFYLLQRTPEVARSRATPLDLIMALTTPHPPOKQAYGVTLPSTVLF 322  
Qy 301 AGHDTNLANLGGALELNWTLPGQDNTPPGGELVFERWRRLSDNSQWIOVSLVFTLQOM 360  
Db 323 AGHDTNLANLGGALELNWTLPGQDNTPPGGELVFERWRRLSDNSQWIOVSLVFTLQOM 382  
Qy 361 RDKTPLSLNTPPGVRLTLGACERNAQGMCSLAGFTQIVNEARIPACSL 410  
Db 383 RDKTPLSLNTPPGVRLTLGACERNAQGMCSLAGFTQIVNEARIPACSL 432

RESULT 3

US-09-291-931-2  
; Sequence 2, Application US/09291931A  
; Patent No. 6130897  
; GENERAL INFORMATION:  
; APPLICANT: Kretz, Keith  
; TITLE OF INVENTION: NOVEL PHYTASE  
; FILE REFERENCE: 09010/029003  
; CURRENT APPLICATION NUMBER: US/09/291,931A  
; CURRENT FILING DATE: 1999-04-13  
; EARLIER APPLICATION NUMBER: 08/910,798  
; EARLIER FILING DATE: 1997-08-13  
; EARLIER APPLICATION NUMBER: 09/259,214

; EARLIER FILING DATE: 1999-03-01  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 440  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-291-931-2

Query Match 96.0%; Score 2066; DB 3; Length 440;  
Best Local Similarity 97.6%; Pred. No. 2.3e-211;  
Matches 400; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 1 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTTPDAPWTPVKLGELTPRGELIAYLGH 60  
Db 23 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTTPDAPWTPVKLGELTPRGELIAYLGH 82  
Qy 61 YMRQRLVADGLLPKCGCPOSGOVAIADVDETRTKTGEAFAAGLAPDCAITVHTQADTSS 120  
Db 83 YQRLVADGLLAKKCGCPOSGOVAIADVDETRTKTGEAFAAGLAPDCAITVHTQADTSS 142  
Qy 121 PDPLFNPLKTVGCOLDNANVTDAILLERAGSGIADFTGHYQTAFARELERVLPQSNCLCK 180  
Db 143 PDPLFNPLKTVGCOLDNANVTDAILLERAGSGIADFTGHYQTAFARELERVLPQSNCLCK 202  
Qy 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 240  
Db 203 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 262  
Qy 241 HOWNTLISLHNAQFDLLORTPEVARSRATPLDLIKTALTTPHPPOKQAYGVTLPSTVLF 300  
Db 263 HOWNTLISLHNAQFYLLQRTPEVARSRATPLDLIMALTTPHPPOKQAYGVTLPSTVLF 322  
Qy 301 AGHDTNLANLGGALELNWTLPGQDNTPPGGELVFERWRRLSDNSQWIOVSLVFTLQOM 360  
Db 323 AGHDTNLANLGGALELNWTLPGQDNTPPGGELVFERWRRLSDNSQWIOVSLVFTLQOM 382  
Qy 361 RDKTPLSLNTPPGVRLTLGACERNAQGMCSLAGFTQIVNEARIPACSL 410  
Db 383 RDKTPLSLNTPPGVRLTLGACERNAQGMCSLAGFTQIVNEARIPACSL 432

RESULT 4

US-09-540-149A-1  
; Sequence 1, Application US/09540149A  
; Patent No. 6511699  
; GENERAL INFORMATION:  
; APPLICANT: Lei, Xingen  
; TITLE OF INVENTION: ENZYMES WITH IMPROVED PHYTASE ACTIVITY  
; FILE REFERENCE: 19603/2791  
; CURRENT APPLICATION NUMBER: US/09/540,149A  
; CURRENT FILING DATE: 2000-03-31  
; PRIOR APPLICATION NUMBER: 60/127,032  
; PRIOR FILING DATE: 1999-03-31  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 433  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (433)  
; OTHER INFORMATION: Xaa at position 433 in this sequence is unknown  
US-09-540-149A-1

Query Match 94.8%; Score 2042; DB 4; Length 433;  
Best Local Similarity 96.6%; Pred. No. 8.1e-209;  
Matches 396; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 1 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTTPDAPWTPVKLGELTPRGELIAYLGH 60  
Db 23 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTTPDAPWTPVKLGELTPRGELIAYLGH 82

QY 61 YWRQRLVADGLLPKCGPQSGQVAIIADVDERTKRTGGEAFAGLAPDCAITVHTQADTSS 120  
DB 83 YQORLQVADGLLAKKCGPQSGQVAIIADVDERTKRTGGEAFAGLAPDCAITVHTQADTSS 142  
QY 121 PDPLFNPLKTVGCOLDNANVTDAIIRAGGSIADFTGHYQTAFARELERVLNFPQSNCLK 180  
DB 143 PDPLFNPLKTVGCOLDNANVTDAIIRAGGSIADFTGHYQTAFARELERVLNFPQSNCLK 202  
QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 240  
DB 203 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 262  
QY 241 HQWNTLLSLHNAQFDLLQRTPEVARSRAATPLDLIKTALTTPHPPOKQAYGVTLPTSVLFI 300  
DB 263 HQWNTLLSLHNAQFYLLQRTPEVARSRAATPLDLIKTALTTPHPPOKQAYGVTLPTSVLFI 322  
QY 301 AGHDTNLANLGGALELNTLPGQDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 360  
DB 323 AGHDTNLANLGGALELNTLPGQDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 382  
QY 361 RDKTPLSLNTPPGEVKLTLAGCBERNAGQCSLAGFTQIVNEARIPACSL 410  
DB 383 RDKTPLSLNTPPGEVKLTLAGCBERNAGQCSLAGFTQIVNEARIPACSL 432

RESULT 5  
US-08-910-798-2  
; Sequence 2, Application US/08910798  
; Patent No. 5876997  
; GENERAL INFORMATION:  
; APPLICANT: KRETZ  
; TITLE OF INVENTION: NOVEL PHYTASE  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: California  
; COUNTRY: US  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/910,798  
; FILING DATE: August 13, 1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: HAILLE, PH.D., LISA A.  
; REGISTRATION NUMBER: 38,347  
; REFERENCE/DOCKET NUMBER: 09010/029001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619/678-5070  
; TELEFAX: 619/678-5099  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 423 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-910-798-2

Query Match 91.0%; Score 1958.5; DB 2; Length 423;  
Best Local Similarity 93.4%; Pred. No. 6.2e-200;  
Matches 383; Conservative 0; Mismatches 10; Indels 17; Gaps 1;  
QY 1 QSEPELKLESVIVSRHGVRAPTKATQLMQDVTDPDAMPTVPVKLGELTPRGGLIAYLGH 60  
DB 23 QSEPELKLESVIVSRHGVRAPTKATQLMQDVTDPDAMPTVPVKLGELTPRGGLIAYLGH 82

QY 61 YWRQRLVADGLLPKCGPQSGQVAIIADVDERTKRTGGEAFAGLAPDCAITVHTQADTSS 120  
DB 83 YQORLQVADGLLAKKCGPQSGQVAIIADVDERTKRTGGEAFAGLAPDCAITVHTQADTSS 125  
QY 121 PDPLFNPLKTVGCOLDNANVTDAIIRAGGSIADFTGHYQTAFARELERVLNFPQSNCLK 180  
DB 126 PDPLFNPLKTVGCOLDNANVTDAIIRAGGSIADFTGHYQTAFARELERVLNFPQSNCLK 185  
QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 240  
DB 186 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 245  
QY 241 HQWNTLLSLHNAQFDLLQRTPEVARSRAATPLDLIKTALTTPHPPOKQAYGVTLPTSVLFI 300  
DB 246 HQWNTLLSLHNAQFYLLQRTPEVARSRAATPLDLIKTALTTPHPPOKQAYGVTLPTSVLFI 305  
QY 301 AGHDTNLANLGGALELNTLPGQDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 360  
DB 306 AGHDTNLANLGGALELNTLPGQDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 365  
QY 361 RDKTPLSLNTPPGEVKLTLAGCBERNAGQCSLAGFTQIVNEARIPACSL 410  
DB 366 RDKTPLSLNTPPGEVKLTLAGCBERNAGQCSLAGFTQIVNEARIPACSL 415

RESULT 6  
US-09-489-039A-7512  
; Sequence 7512, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 7512  
; LENGTH: 421  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-7512

Query Match 25.0%; Score 539; DB 4; Length 421;  
Best Local Similarity 32.3%; Pred. No. 1.2e-48;  
Matches 132; Conservative 70; Mismatches 181; Indels 26; Gaps 9;  
QY 1 QSEPE-LKLESVIVSRHGVRAPTKATQLMQDVTDPDAMPTVPVKLGELTPRGGLIAYL 58  
DB 31 KAAPEGYQLQVLLMSRHLRAPLANNGSVLEQSTAKAWPDVGGGLTTKGGVLEYVM 90  
QY 59 GHYWRQLVADGLLPKCGPQSGQVAIIADVDERTKRTGGEAFAGLAPDCAITVHTQADT 118  
DB 91 GHYWRQLVADGLLPKCGPQSGQVAIIADVDERTKRTGGEAFAGLAPDCAITVHTQADT 150  
QY 119 SSPDPLFNPLKTVGCOLDNANVTDAI-IERAGGSIADFTGHYQTAFARELERVLNFPQSNL 177  
DB 151 GTMDPTFNFVITDSDPAFPAKALQAMEKEROQMTE-----SYKLETFMIDIRNPS 203  
QY 178 CLKREKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPE--EPGWG 235  
DB 204 C---KEKVCVLSSEKDTFSGYQEQEPGVSGPLKVGNSLVDVDAFTLQYVEGPKDQVANG 259  
QY 236 RITDQWNTLLSLHNAQFDLLQRTPEVARSRAATPLDLIKTALTTPHPPOKQAYGVTLPT 295  
DB 260 EIADKQWRVLSKLNKNGYQDSLFTSVAQVAKVLYKIDNALYVEGASK-----A 311  
QY 296 SVLFIAGHDTNLANLGGALELNTLPGQDNTPPGGELVFERWRRLSDNSQWIOVSLVF 354  
DB 312 KVTLLVGHDSNTASLTLTALDFFPYQLPGQYERTPIGGKLLFORWHDNSAGNRDLMEIYVY 371



Query Match 5.5%; Score 118.5; DB 4; Length 439;  
Best Local Similarity 20.9%; Pred. No. 0.00087;  
Matches 100; Conservative 65; Mismatches 157;  
Indels 157; Gaps 26;

RESULT 8  
US-09-684-855-165  
; Sequence 165, Application US/09684855  
; Patent No. 6599735  
; GENERAL INFORMATION:  
; APPLICANT: F. Hoffmann-La Roche AG  
; TITLE OF INVENTION: CONTINUOUS FERMENTATION PROCESS  
; FILE REFERENCE: C38435/111692  
; CURRENT APPLICATION NUMBER: US/09/684,855  
; CURRENT FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: EP 00121663.9  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: EP 99120289.6  
; PRIOR FILING DATE: 1999-10-11  
; NUMBER OF SEQ ID NOS: 169

RESULT 8  
US-09-684-855-165  
; Sequence 165, Application US/09684855  
; Patent No. 6599735  
; GENERAL INFORMATION:  
; APPLICANT: F. Hoffmann-La Roche AG  
; TITLE OF INVENTION: CONTINUOUS FERMENTATION PROCESS  
; FILE REFERENCE: C38435/111692  
; CURRENT APPLICATION NUMBER: US/09/684,855  
; CURRENT FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: EP 00121663.9  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: EP 99120289.6  
; PRIOR FILING DATE: 1999-10-11  
; NUMBER OF SEQ ID NOS: 169

QY 3 EPELKLES-----VVIVSRHGVRAPT-----KATQLMQDVTTPDA-----WP 38  
DB 31 EDELSVSSKLPKDCRITLVQVLSRHGARYPTSSKSKYKLVTAIQANATDFKGFALFK 90  
QY 39 TWPVKLG--ELTPRGELIAYLG--HYWRQRLVADGLLPKCGCPSQGVVAIIADVDETR 94  
DB 91 TYNVTLGADDLTPFGEOQLVNSGKIFYORYKALARSVVP-----FIRASGSDRVI 140  
QY 95 KTGEAFAAGL-----APDCAITVHTQADTSSPDPLFNPLKTVGC-OLDNAN 139  
DB 141 ASGEKFIQFQOAKLADPGATNRAAPAIISVII-PESETFN-----NTLDHGVCITKFEASQ 194  
QY 140 VTDAILERAGGSIAFTGHYQTAFARELVRVLPFQSNCLKREKQDESCSLTQALPSELK 199  
DB 195 LGDEV-----AANFTALFAPDIR-----ARAEKH-----LPGVTL 224  
QY 200 VSADCVSLTGAIVSLASMLTEIFLLQQAQMPFGWGRITDSHQWNTLLSL--HN--AQFD 255  
DB 225 TDEDVVSMDMCSFDTV-----ARTSDASQLSPFCQLFTHNEWKYN 266  
QY 256 LLQR-----TPE-VARSRATPLLDLIKT--ALTPHPQKQAYGV 291  
DB 267 YLQSLGKYGYGAGNPLGPAQIGFTNELIARLTRSPVQDHTSTNSTLVSNP-----A 319  
QY 292 TLPTSVLFTA--GHDNTNLNGLGALNLTLPQPDNTPPGGELVFERWRRLSD-NSOWI 348  
DB 320 TFFLNATMYVDFSHDMSVSIFFAL-----GLYNGTEPLSRTSVESAKELDGYASWV 372  
QY 349 ---QVSLVFOTLQOMRDKTPLSLNTPPGEVKLTLAGCEERNAQMGCSLAGFTQIVNEAR 404  
DB 373 VPFGARAYFETMCKSEKEPL-VRALINDRVVPLHGCDV-DKLGRCCKLNDVFKGLSWAR 429  
RESULT 10  
US-09-684-855-106  
; Sequence 106, Application US/09684855  
; Patent No. 6599735  
; GENERAL INFORMATION:  
; APPLICANT: F. Hoffmann-La Roche AG  
; TITLE OF INVENTION: CONTINUOUS FERMENTATION PROCESS  
; FILE REFERENCE: C38435/111692  
; CURRENT FILING DATE: 2000-10-06  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR FILING DATE: 1999-10-11  
; NUMBER OF SEQ ID NOS: 169  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 106  
; LENGTH: 440  
; TYPE: PRT  
; ORGANISM: A. fumigatus 13073  
US-09-684-855-106

Query Match 5.5%; Score 118.5; DB 4; Length 440;  
Best Local Similarity 20.9%; Pred. No. 0.00088;  
Matches 100; Conservative 65; Mismatches 157; Indels 157; Gaps 26;  
QY 3 EPELKLES-----VVIVSRHGVRAPT-----KATQLMQDVTTPDA-----WP 38  
DB 32 EDELSVSSKLPKDCRITLVQVLSRHGARYPTSSKSKYKLVTAIQANATDFKGFALFK 91  
QY 39 TWPVKLG--ELTPRGELIAYLG--HYWRQRLVADGLLPKCGCPSQGVVAIIADVDETR 94  
DB 92 TYNVTLGADDLTPFGEOQLVNSGKIFYORYKALARSVVP-----FIRASGSDRVI 141  
QY 95 KTGEAFAAGL-----APDCAITVHTQADTSSPDPLFNPLKTVGC-OLDNAN 139  
DB 142 ASGEKFIQFQOAKLADPGATNRAAPAIISVII-PESETFN-----NTLDHGVCITKFEASQ 195  
QY 140 VTDAILERAGGSIAFTGHYQTAFARELVRVLPFQSNCLKREKQDESCSLTQALPSELK 199  
DB 195 LGDEV-----AANFTALFAPDIR-----ARAEKH-----LPGVTL 225  
QY 200 VSADCVSLTGAIVSLASMLTEIFLLQQAQMPFGWGRITDSHQWNTLLSL--HN--AQFD 255  
DB 225 TDEDVVSMDMCSFDTV-----ARTSDASQLSPFCQLFTHNEWKYN 267  
QY 256 LLQR-----TPE-VARSRATPLLDLIKT--ALTPHPQKQAYGV 291  
DB 267 YLQSLGKYGYGAGNPLGPAQIGFTNELIARLTRSPVQDHTSTNSTLVSNP-----A 320  
QY 292 TLPTSVLFTA--GHDNTNLNGLGALNLTLPQPDNTPPGGELVFERWRRLSD-NSOWI 348  
DB 320 TFFLNATMYVDFSHDMSVSIFFAL-----GLYNGTEPLSRTSVESAKELDGYASWV 373  
QY 349 ---QVSLVFOTLQOMRDKTPLSLNTPPGEVKLTLAGCEERNAQMGCSLAGFTQIVNEAR 404  
DB 373 VPFGARAYFETMCKSEKEPL-VRALINDRVVPLHGCDV-DKLGRCCKLNDVFKGLSWAR 429

DB 196 LGDEV-----AANFTALFAPDIR-----ARAEKH-----LPGVTL 225  
QY 200 VSADCVSLTGAIVSLASMLTEIFLLQQAQMPFGWGRITDSHQWNTLLSL--HN--AQFD 255  
DB 226 TDEDVVSMDMCSFDTV-----ARTSDASQLSPFCQLFTHNEWKYN 267  
QY 256 LLQR-----TPE-VARSRATPLLDLIKT--ALTPHPQKQAYGV 291  
DB 268 YLQSLGKYGYGAGNPLGPAQIGFTNELIARLTRSPVQDHTSTNSTLVSNP-----A 320  
QY 292 TLPTSVLFTA--GHDNTNLNGLGALNLTLPQPDNTPPGGELVFERWRRLSD-NSOWI 348  
DB 321 TFFLNATMYVDFSHDMSVSIFFAL-----GLYNGTEPLSRTSVESAKELDGYASWV 373  
QY 349 ---QVSLVFOTLQOMRDKTPLSLNTPPGEVKLTLAGCEERNAQMGCSLAGFTQIVNEAR 404  
DB 374 VPFGARAYFETMCKSEKEPL-VRALINDRVVPLHGCDV-DKLGRCCKLNDVFKGLSWAR 430  
RESULT 11  
US-09-684-855-128  
; Sequence 128, Application US/09684855  
; Patent No. 6599735  
; GENERAL INFORMATION:  
; APPLICANT: F. Hoffmann-La Roche AG  
; TITLE OF INVENTION: CONTINUOUS FERMENTATION PROCESS  
; FILE REFERENCE: C38435/111692  
; CURRENT FILING DATE: 2000-10-06  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR FILING DATE: 1999-10-11  
; NUMBER OF SEQ ID NOS: 169  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 128  
; LENGTH: 440  
; TYPE: PRT  
; ORGANISM: A. fumigatus 13073  
US-09-684-855-128

Query Match 5.5%; Score 118.5; DB 4; Length 440;  
Best Local Similarity 20.9%; Pred. No. 0.00088;  
Matches 100; Conservative 65; Mismatches 157; Indels 157; Gaps 26;  
QY 3 EPELKLES-----VVIVSRHGVRAPT-----KATQLMQDVTTPDA-----WP 38  
DB 32 EDELSVSSKLPKDCRITLVQVLSRHGARYPTSSKSKYKLVTAIQANATDFKGFALFK 91  
QY 39 TWPVKLG--ELTPRGELIAYLG--HYWRQRLVADGLLPKCGCPSQGVVAIIADVDETR 94  
DB 92 TYNVTLGADDLTPFGEOQLVNSGKIFYORYKALARSVVP-----FIRASGSDRVI 141  
QY 95 KTGEAFAAGL-----APDCAITVHTQADTSSPDPLFNPLKTVGC-OLDNAN 139  
DB 142 ASGEKFIQFQOAKLADPGATNRAAPAIISVII-PESETFN-----NTLDHGVCITKFEASQ 195  
QY 140 VTDAILERAGGSIAFTGHYQTAFARELVRVLPFQSNCLKREKQDESCSLTQALPSELK 199  
DB 196 LGDEV-----AANFTALFAPDIR-----ARAEKH-----LPGVTL 225  
QY 200 VSADCVSLTGAIVSLASMLTEIFLLQQAQMPFGWGRITDSHQWNTLLSL--HN--AQFD 255  
DB 226 TDEDVVSMDMCSFDTV-----ARTSDASQLSPFCQLFTHNEWKYN 267  
QY 256 LLQR-----TPE-VARSRATPLLDLIKT--ALTPHPQKQAYGV 291  
DB 268 YLQSLGKYGYGAGNPLGPAQIGFTNELIARLTRSPVQDHTSTNSTLVSNP-----A 320  
QY 292 TLPTSVLFTA--GHDNTNLNGLGALNLTLPQPDNTPPGGELVFERWRRLSD-NSOWI 348  
DB 321 TFFLNATMYVDFSHDMSVSIFFAL-----GLYNGTEPLSRTSVESAKELDGYASWV 373

Qy 349 ---QVSLVFOTLQMRDKTPLSLNTPPEVKLTLAGCEERNAQMGSLAGFTQIVNEAR 404  
 Db 374 VPFGARAYFETMCKSEKEPL-VRALINDRVVPLHGCDV-DKLGRCCKLNDVFKGLSWAR 430

RESULT 12

US-09-684-855-151  
 ; Sequence 151, Application US/09684855  
 ; Patent No. 6599735  
 ; GENERAL INFORMATION:  
 ; APPLICANT: F. Hoffmann-La Roche AG  
 ; TITLE OF INVENTION: CONTINUOUS FERMENTATION PROCESS  
 ; FILE REFERENCE: C38435/111692  
 ; CURRENT APPLICATION NUMBER: US/09/684,855  
 ; PRIOR FILING DATE: 2000-10-06  
 ; PRIOR APPLICATION NUMBER: EP 00121663.9  
 ; PRIOR FILING DATE: 2000-10-04  
 ; PRIOR APPLICATION NUMBER: EP 99120289.6  
 ; PRIOR FILING DATE: 1999-10-11  
 ; NUMBER OF SEQ ID NOS: 169  
 ; SOFTWARE: Patent in Ver. 2.0  
 ; SEQ ID NO 151  
 ; LENGTH: 440  
 ; TYPE: PRT  
 ; ORGANISM: A. fumigatus 32722  
 ; US-09-684-855-151

Query Match 5.5%; Score 118.5; DB 4; Length 440;  
 Best Local Similarity 20.9%; Pred. No. 0.00088;  
 Matches 100; Conservative 65; Mismatches 157; Indels 157; Gaps 26;

Qy 3 EPELKLES-----VVIVSRHGVRAPT-----KATQIMODVTPDA-----WP 38  
 Db 32 EDELSVSKLPKDCRITLVQVLSRHGARYPTSSKSKYKLVTAIQANATDFKGFAPLK 91  
 Qy 39 TWPKVIG--ELTPRGGLIAYLG--HYWRORLVADGLPKCGCPQSGQVAILIADVDERT 94  
 Db 92 TYNVTLGADDLTPFGEQQLVNSGKIFQRYKALARSVVP-----FIRASGSDRVI 141  
 Qy 95 KTGEAFAGL-----APDCAITVHTQADTSSPDLPFLNPLKTGVC--QLDNAN 139  
 Db 142 ASGEKFTGEGQOAKLADPGATNRAAPASVII--PESETFN-----NTLDHGVCCTKFEASQ 195  
 Qy 140 VTDAILERAGGSIAIDFTGHYQTAFRELERVINFQSNCLKREKQDESCSLTQALPSELK 199  
 Db 196 LGDEV-----AANFTALFAPDIR-----ARAEKH-----LPGVTL 225  
 Qy 200 VSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRIITDSHOWNTLLSL--HN--AQFD 255  
 Db 226 TDEDVWSLMDMCSFDTV-----ARTSDASQLSPFCQLFTHNEWKYN 267  
 Qy 256 LLQR-----TPE-VARSRATPLLDLIKT--ALTPHPPOKQAYGV 291  
 Db 268 YLOSGLKYYGYGAGNPLGPAQGGIFTNELIARLTRSPVQDHTSTNSTLVSNP-----A 320  
 Qy 292 TLPTSULFIA--GHDNTNLANLGALBNWTLPGOPDNTPPGGELVFRWRRLSD--NSQWI 348  
 Db 321 TFFLNATMYVDFSHDMSVSIFFAL-----GLYNGTEPLSRTSVESAKELDGYASVW 373  
 Qy 349 ---QVSLVFOTLQMRDKTPLSLNTPPEVKLTLAGCEERNAQMGSLAGFTQIVNEAR 404  
 Db 374 VPFGARAYFETMCKSEKEPL-VRALINDRVVPLHGCDV-DKLGRCCKLNDVFKGLSWAR 430

RESULT 13

US-09-044-718-12  
 ; Sequence 12, Application US/09044718  
 ; Patent No. 6391605  
 ; GENERAL INFORMATION:  
 ; APPLICANT: KOSTREWA, Dirk  
 ; APPLICANT: PASAMONTES, Luis  
 ; APPLICANT: TOMSCHY, Andrea  
 ; APPLICANT: van LOON, Adolphus

; APPLICANT: VOGEL, Kurt  
 ; APPLICANT: WYSS, Markus  
 ; TITLE OF INVENTION: MODIFIED PHYTASES  
 ; FILE REFERENCE: Modified Phytases  
 ; CURRENT APPLICATION NUMBER: US/09/044,718  
 ; PRIOR FILING DATE: 1998-03-19  
 ; PRIOR APPLICATION NUMBER: EP 97810175.6  
 ; PRIOR FILING DATE: 1997-03-25  
 ; NUMBER OF SEQ ID NOS: 82  
 ; SOFTWARE: Patent in Ver. 2.1  
 ; SEQ ID NO 12  
 ; LENGTH: 449  
 ; TYPE: PRT  
 ; ORGANISM: Aspergillus fumigatus  
 ; US-09-044-718-12

Query Match 5.5%; Score 118.5; DB 4; Length 449;  
 Best Local Similarity 20.9%; Pred. No. 0.00091;  
 Matches 100; Conservative 65; Mismatches 157; Indels 157; Gaps 26;

Qy 3 EPELKLES-----VVIVSRHGVRAPT-----KATQIMODVTPDA-----WP 38  
 Db 41 EDELSVSKLPKDCRITLVQVLSRHGARYPTSSKSKYKLVTAIQANATDFKGFAPLK 100  
 Qy 39 TWPKVIG--ELTPRGGLIAYLG--HYWRORLVADGLPKCGCPQSGQVAILIADVDERT 94  
 Db 101 TYNVTLGADDLTPFGEQQLVNSGKIFQRYKALARSVVP-----FIRASGSDRVI 150  
 Qy 95 KTGEAFAGL-----APDCAITVHTQADTSSPDLPFLNPLKTGVC--QLDNAN 139  
 Db 151 ASGEKFTGEGQOAKLADPGATNRAAPASVII--PESETFN-----NTLDHGVCCTKFEASQ 204  
 Qy 140 VTDAILERAGGSIAIDFTGHYQTAFRELERVINFQSNCLKREKQDESCSLTQALPSELK 199  
 Db 205 LGDEV-----AANFTALFAPDIR-----ARAEKH-----LPGVTL 234  
 Qy 200 VSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRIITDSHOWNTLLSL--HN--AQFD 255  
 Db 235 TDEDVWSLMDMCSFDTV-----ARTSDASQLSPFCQLFTHNEWKYN 276  
 Qy 256 LLQR-----TPE-VARSRATPLLDLIKT--ALTPHPPOKQAYGV 291  
 Db 277 YLOSGLKYYGYGAGNPLGPAQGGIFTNELIARLTRSPVQDHTSTNSTLVSNP-----A 329  
 Qy 292 TLPTSULFIA--GHDNTNLANLGALBNWTLPGOPDNTPPGGELVFRWRRLSD--NSQWI 348  
 Db 330 TFFLNATMYVDFSHDMSVSIFFAL-----GLYNGTEPLSRTSVESAKELDGYASVW 382  
 Qy 349 ---QVSLVFOTLQMRDKTPLSLNTPPEVKLTLAGCEERNAQMGSLAGFTQIVNEAR 404  
 Db 383 VPFGARAYFETMCKSEKEPL-VRALINDRVVPLHGCDV-DKLGRCCKLNDVFKGLSWAR 439

RESULT 14

US-08-868-435-33  
 ; Sequence 33, Application US/08868435  
 ; Patent No. 6291221  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Van Loon, Adolphus  
 ; APPLICANT: Mitchell, David  
 ; TITLE OF INVENTION: POLYPEPTIDES WITH PHYTASE ACTIVITY  
 ; NUMBER OF SEQUENCES: 35  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Hoffmann-La Roche Inc.  
 ; STREET: 340 Kingsland Street  
 ; CITY: Nutley  
 ; STATE: New Jersey  
 ; COUNTRY: United States of America  
 ; ZIP: 07110  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/868,435  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA: 08/744,231  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Kass, Alan P  
REGISTRATION NUMBER: 32142  
REFERENCE/DOCKET NUMBER: Case Docket 9339  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (201) 235-4205  
TELEFAX: (201) 235-2363  
INFORMATION FOR SEQ ID NO: 33:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 465 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 104  
OTHER INFORMATION: /note="potential N-glycosylation site"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 119  
OTHER INFORMATION: /note="potential N-glycosylation site"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 205  
OTHER INFORMATION: /note="potential N-glycosylation site"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 228  
OTHER INFORMATION: /note="potential N-glycosylation site"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 337  
OTHER INFORMATION: /note="potential N-glycosylation site"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 374  
OTHER INFORMATION: /note="potential N-glycosylation site"  
US-08-868-435-33  
Query Match 5.5%; Score 118.5; DB 3; Length 465;  
Best Local Similarity 20.9%; Pred. No. 0.00096;  
Matches 100; Conservative 65; Mismatches 157; Indels 157; Gaps 26;  
Qy 3 EPELKLES-----VVIVSRHGVRAPT-----KATQLMDQVTPDA-----WP 38  
Db 57 EDELSVSKLPKDCRITLVQLVSRHGARYPTSSKSKYKLVTAIQANATDFKGFAPLK 116  
Qy 39 TWPVKLG--ELTPRGGLIAYLG--HWYRORLVADGLLPKCGCPQSQVAAIADVERTR 94  
Db 117 TYNLTGADDLTPFGEOQLVNSGKIFYQRYKALARSVP-----FIRASGSDRVI 166  
Qy 95 KTGFAFAAGL-----APCAITVHTQADTSSPDPLFNPLKTVGC--OLDNAN 139  
Db 167 ASGKEPIEGFOQAKLADPGATNRAAPALSVII--PESETFN-----NVLDHGVCTKFEASQ 220  
Qy 140 VTDAILERAGGSIADFTGHVQTAFRELERVLPQSNLCUKREKQDECSLTQALPSELK 199  
Db 221 LGDEV-----AANFTALFAPDIR-----ARAERH-----LPGVTL 250  
Qy 200 VSACVSVLTGAVSLASMLTEIFLLQQAQMGPEPCMGRIITSHOWNTLLSL--HN--AQFD 255  
Db 251 TDEDVSLMDMCSFDTV-----ARTSDASQLSPFCQLFTHEWKKYN 292  
Qy 256 LLQR-----TPE-VARSRATPLLDLIKT--ALTPHPQKQAYGV 291

Db 293 YLOSLGKYGYGAGNPLGPAQIGFTNELLARTSPVQDHTSTNSTLVNRP-----A 345  
Qy 292 TLPTSVLFFIA--GHDITNLANLGGALELNWTLPGQPDNTPPGGELVFERWRLSD--NSQWI 348  
Db 346 TFFLNATMYVDFSHDMSVSIFFAL-----GLYNGTEPLSRTSVESAKELDGYASAWV 398  
Qy 349 ---QVSLVFQTLQOMRDKTPLSLNTPFGEVKKLTLAGCEERNAQGMCSLAGFTQIVNEAR 404  
Db 399 VPFGARAYFETMOCKSEKPL--VRALINDRVVPLHGCDV--DKLGRCKINDFVKGLSWAR 455  
RESULT 15  
US-08-744-231-33  
Sequence 33, Application US/08744231  
Patent No. 6358722  
GENERAL INFORMATION:  
APPLICANT: Van Loon, Adolphus  
APPLICANT: Mitchell, David  
TITLE OF INVENTION: POLYPEPTIDES WITH PHYTASE ACTIVITY  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hoffmann-La Roche Inc.  
STREET: 340 Kingsland Street  
CITY: Nutley  
STATE: New Jersey  
COUNTRY: United States of America  
ZIP: 07110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/744,231  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/424,757  
FILING DATE: 18-APR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Kass, Alan P  
REGISTRATION NUMBER: 32142  
REFERENCE/DOCKET NUMBER: Case Docket 9339  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (201) 235-4205  
TELEFAX: (201) 235-2363  
INFORMATION FOR SEQ ID NO: 33:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 465 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 104  
OTHER INFORMATION: /note="potential N-glycosylation site"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 119  
OTHER INFORMATION: /note="potential N-glycosylation site"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 205  
OTHER INFORMATION: /note="potential N-glycosylation site"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 228  
OTHER INFORMATION: /note="potential N-glycosylation site"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 337  
OTHER INFORMATION: /note="potential N-glycosylation site"  
FEATURE:

```
; NAME/KEY: misc_feature
; LOCATION: 374
; OTHER INFORMATION: /note="potential N-glycosylation site"
US-08-744-231-33

Query Match      5.5%; Score 118.5; DB 4; Length 465;
Best Local Similarity 20.9%; Pred. No. 0.00096;
Matches 100; Conservative 65; Mismatches 157; Indels 157; Gaps 26;

QY 3 EPELKLES-----VVIVSRHGVRAPT-----KATQLMQDVTTPDA-----WP 38
Db 57 EDELSVSSKLPKDCRITLVQVLSRHGARYPTSSKKYKLVTAIQANATDFKGFAPLK 116
QY 39 TWPVKLG--ELTPRGGELIAYLG--HYWRQRLVADGILLPKCGPQSGQVAILIADVDERTR 94
Db 117 TYNVTLGADDLTFGQQLVNSGIKFYQYKALARSVP-----FIRASGSDRAVI 166
QY 95 KTGEAFAGL-----APDCAITVHTQADTSSDPDLFNPPLKTGVC-QLDNAN 139
Db 167 ASGEKFIEGFQQAQLADPGATNRAAPAIISVII-PESETFN-----NTLDHGVCTKFEASQ 220
QY 140 VTDAILERAGGSIADEFTGHYQTAFRELERVLNFPQSNLCLKREKQDESCSLTQALPSELK 199
Db 221 LGDEV-----AANTALFAPDIR-----ARAEKH-----LFGVTL 250
QY 200 VSADCVSLTGAVSLASMLTEIFILQQAQGMPEPGWGRITDSHOWNTLLSL--HN--AQPD 255
Db 251 TDEDVVSMDMCSFDTV-----ARTSDASQLSPFCQLFTHNEWKKYN 292
QY 256 LLQR-----TPE-VARSRAATPLDLIKT--ALTPHPQKQAYGV 291
Db 293 YLOSGLKYYGYGAGNPLGPAQGIPTNELIARLTRSPVDHTSTNSTLVSNP-----A 345
QY 292 TLPTSVLFTA--GHDNLANLWLGALINWTLPGQPDNTPPGGELVFERWRRLSD-NSQWI 348
Db 346 TFEPLNATMYVDFSHDNSMWSIFFAL-----GLYNGTEPLSRTSVESAKELDGYASAV 398
QY 349 ---QVSLVFOTLQOMRDKTPLSLNTPPGEVKLTLAGCEERNAQCMCSLAGFTQIVNEAR 404
Db 399 VPEGARAYPETMQCKSEKEPL-VRALINDRVPLHGCDV-DKLGRCCKLNDFVKGLSWAR 455
```

Search completed: April 30, 2004, 12:43:17  
Job time : 17.4286 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 12:30:35 ; Search time 52.5 Seconds

(without alignments)  
2206.561 Million cell updates/sec

Title: US-10-021-723B-13

Perfect score: 2153

Sequence: 1 QSEPEKLSEVIVSRHGVR.....CSLAGFTQVNEARIPACSL 410

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_29Jan04:\*  
1: Geneseq1980s:\*  
2: Geneseq1990s:\*  
3: Geneseq2000s:\*  
4: Geneseq2001s:\*  
5: Geneseq2002s:\*  
6: Geneseq2003as:\*  
7: Geneseq2003bs:\*  
8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2153	100.0	410	5	ABP51937 Phytase p
2	2153	100.0	412	6	AAE37851 Maize-opt
3	2153	100.0	412	7	ABR82310 Nov9x phy
4	2153	100.0	431	6	AAE37853 pNOV4054
5	2153	100.0	436	7	ADC87743 Modified
6	2153	100.0	437	6	AAE37854 pNOV4058
7	2123	98.6	430	5	AAE15808 Escherich
8	2123	98.6	430	6	ADA19451 E. coli K
9	2077	96.5	432	3	AAE36257 Lama2/APP
10	2077	96.5	432	3	AAE36261 R15/APP
11	2077	96.5	432	3	AAE36262 SV40/APP
12	2077	96.5	432	3	AAE36259 R15/APP
13	2077	96.5	432	3	AAE36258 R15/APP
14	2077	96.5	432	3	AAE36263 Lama2/APP
15	2077	96.5	432	3	AAU7775 Phytase a
16	2077	96.5	432	5	AAE15807 Escherich
17	2077	96.5	432	5	ABP3625 Acid phos
18	2077	96.5	432	6	ADA19450 E. coli K
19	2077	96.5	432	6	ABU15427 Protein e
20	2077	96.5	432	7	ADC87745 Escherich
21	2077	96.5	432	4	AAE02631 E. coli a
22	2072	96.2	432	4	AAE02634 E. coli a
23	2066	96.0	440	4	AAE37892 Escherich
24	2066	96.0	440	5	AAE22836 Escherich
25	2066	96.0	440	5	AAE22836 Escherich

26	2066	96.0	440	5	AAE15806
27	2066	96.0	440	6	ADA19446
28	2065	95.9	432	4	AAE02635
29	2052	95.3	432	4	AAE02632 E. coli a
30	2042	94.8	432	3	AAV94753
31	2042	94.8	432	6	ABR42161
32	2039	94.7	432	6	ABR42162
33	1958.5	91.0	423	2	AAV01513
34	1192	55.4	261	4	ABG25698
35	1121.5	52.1	393	5	ABP51935
36	971	45.1	441	5	ABP51934
37	954	44.3	441	6	ABU49868
38	946	43.9	441	5	ABP51929
39	943	43.8	420	5	ABP51930
40	773.5	35.9	476	5	ABP51933
41	637	29.6	144	4	ABG24256
42	613.5	28.5	318	5	ABP51936
43	483	22.4	123	4	AAU23269
44	463.5	21.5	409	5	ABP51932
45	458.5	21.3	421	5	ABP51931

## ALIGNMENTS

### RESULT 1

ABP51937  
ID ABP51937 standard; protein; 410 AA.

XX AC ABP51937;

XX DT 08-OCT-2002 (first entry)

XX DE Phytase protein sequence.

XX KW Phytase; enzyme; E.C.3.1.3.8; phytate; foodstuff; feed.

XX OS Unidentified.

XX PN WO200248332-A2.

XX PD 20-JUN-2002.

XX PF 12-DEC-2001; 2001WO-US048774.

XX PR 12-DEC-2000; 2000US-0255090P.

XX PA (DIVE-) DIVERSA CORP.

XX PI Short J, Mathur EJ, Richardson T, Robertson D, Barton N;

XX DR WPI; 2002-583504/62.

XX PT Novel recombinant phytase protein and polynucleotide for improving

XX PT nutritional value of phytate-containing foodstuff, in animal feed and

XX PT feed supplements and to degrade excess phytase from environment or

XX PT sample.

XX PS Disclosure; Fig 9; 208pp; English.

The present invention describes an isolated phytase protein (I). (I) can be used for improving the nutritional value of a phytate-containing foodstuff, by contacting the phytate-containing foodstuff with (I), where the enzyme catalyses the liberation of inorganic phosphate from the phytate-containing foodstuff, and so improving the nutritive value of the contacted foodstuff. The liberation of inorganic phosphate occurs prior to or after the ingestion of phytate-containing foodstuff by a recipient organism. Nucleotide sequences (II) encoding (I) can be used for producing an animal feed, by transforming a plant, plant portion or plant cell with a nucleic acid expression vector, comprising (II), culturing the plant, plant portion or plant cell under conditions in which the phytase protein is expressed and converting the plant, plant portion or plant cell into a composition suitable for animal feed. The animal is

CC preferably a monogastric animal or a ruminant. (I) and (II) are useful in  
 CC animal feed and feed supplements as well as in treatments to degrade or  
 CC remove excess phytate from the environment or a sample. (I) reduces  
 CC phytate levels in animal manure and so reduces the phosphate pollution of  
 CC the environment. The present sequence represents a phytase from the  
 CC present invention

XX SQ Sequence 410 AA;  
 Query Match 100.0%; Score 2153; DB 5; Length 410;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-210;  
 Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QSEPELKLESVIVSRHGVRAPTKATQMDVTPDAMPVTKLGGELTAPRGGELIAYLGH 60  
 DB 1 QSEPELKLESVIVSRHGVRAPTKATQMDVTPDAMPVTKLGGELTAPRGGELIAYLGH 60  
 QY 61 YWQRVLVADGLLPKCGCPQSGQVAIIADVDERTKGTGEAFAAGLAPDCAITVHTQADTSS 120  
 DB 61 YWQRVLVADGLLPKCGCPQSGQVAIIADVDERTKGTGEAFAAGLAPDCAITVHTQADTSS 120  
 QY 121 PDPLFNPLKTVGCQOLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLPQSNLCIK 180  
 DB 121 PDPLFNPLKTVGCQOLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLPQSNLCIK 180  
 QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFILQOAGMPEPGWGRITDS 240  
 DB 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFILQOAGMPEPGWGRITDS 240  
 QY 241 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLLDLIKTALTTPHPPOKQAYGVTLPSTVLFI 300  
 DB 241 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLLDLIKTALTTPHPPOKQAYGVTLPSTVLFI 300  
 QY 301 AGHDTNLANLGGALELNTWLPFGPDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 360  
 DB 301 AGHDTNLANLGGALELNTWLPFGPDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 360  
 QY 361 RDKTPLSLNTPPGEVKLTLAGCERNAGMCSLAGFTQIVNEARIPACSL 410  
 DB 361 RDKTPLSLNTPPGEVKLTLAGCERNAGMCSLAGFTQIVNEARIPACSL 410

RESULT 2  
 AAE37851  
 ID AAE37851 standard; protein; 412 AA.  
 XX  
 AC AAE37851;  
 XX  
 DT 06-NOV-2003 (first entry)  
 XX  
 DE Maize-optimised Nov9X phytase mutant.  
 XX  
 KW Thermotolerant phytase; weight gain; animals fed diet; grain processing;  
 XX nutritive value; transformed plant; anabolic; maize; mutant; mutein.  
 XX  
 OS Zea mays.  
 OS Synthetic.  
 XX  
 PN WO2003057248-A1.  
 XX  
 PD 17-JUL-2003.  
 XX  
 PF 30-DEC-2002; 2002WO-US041787.  
 XX  
 PR 28-DEC-2001; 2001US-0344476P.  
 XX  
 PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
 XX  
 PI Lanahan MB, Betts S;  
 XX  
 DR WPI; 2003-607980/57.  
 DR N-PSDB; AAD57147.  
 XX

PT Preparing a thermotolerant phytase for preparing animal feed or human  
 PT food by expressing in a plant cell an expression cassette comprising a  
 PT promoter operably linked to a nucleic acid molecule encoding a  
 XX thermotolerant phytase.

PS Claim 3; Page 89; 157pp; English.

XX  
 CC The invention relates to a method for preparing a thermotolerant phytase.  
 CC The method comprises expressing in a plant cell an expression cassette  
 CC comprising a promoter operably linked to a nucleic acid molecule encoding  
 CC a thermotolerant phytase which retains at least 40% activity after 30  
 CC minutes at 60plusOC and has a specific activity of greater than 200 U/mg  
 CC at pH 4.5 and 3pHusOC. The method is useful for preparing a  
 CC thermotolerant phytase for preparing animal feed or human food. The  
 CC invention is useful for reducing the feed conversion ratio and increasing  
 CC weight gain, improving reducing feed conversion ratios or increasing  
 CC weight gain of animals fed diets with inorganic phosphate at levels below  
 CC 0.45%, minimising dietary requirements of phosphorus in an animal,  
 CC enhancing the utilisation of phosphorus present in animal feed, enhancing  
 CC organic phosphorus utilisation from organic phosphorus sources in animal  
 CC feed, decreasing the phosphate levels in excreta from an animal,  
 CC improving the processing of grain, improving the nutritive value of  
 CC processed grain product or a method of processing grain, improving the  
 CC nutritive value of animal feed and human food, and preparing a  
 CC transformed plant which expresses a thermotolerant phytase. The present  
 CC sequence is maize-optimised Nov9X phytase mutant used in the  
 CC exemplification of the invention

XX Sequence 412 AA;

Query Match 100.0%; Score 2153; DB 6; Length 412;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-210;  
 Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QSEPELKLESVIVSRHGVRAPTKATQMDVTPDAMPVTKLGGELTAPRGGELIAYLGH 60  
 DB 3 QSEPELKLESVIVSRHGVRAPTKATQMDVTPDAMPVTKLGGELTAPRGGELIAYLGH 62  
 QY 61 YWQRVLVADGLLPKCGCPQSGQVAIIADVDERTKGTGEAFAAGLAPDCAITVHTQADTSS 120  
 DB 63 YWQRVLVADGLLPKCGCPQSGQVAIIADVDERTKGTGEAFAAGLAPDCAITVHTQADTSS 122  
 QY 121 PDPLFNPLKTVGCQOLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLPQSNLCIK 180  
 DB 123 PDPLFNPLKTVGCQOLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLPQSNLCIK 182  
 QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFILQOAGMPEPGWGRITDS 240  
 DB 183 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFILQOAGMPEPGWGRITDS 242  
 QY 241 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLLDLIKTALTTPHPPOKQAYGVTLPSTVLFI 300  
 DB 243 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLLDLIKTALTTPHPPOKQAYGVTLPSTVLFI 302  
 QY 301 AGHDTNLANLGGALELNTWLPFGPDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 360  
 DB 303 AGHDTNLANLGGALELNTWLPFGPDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 362  
 QY 361 RDKTPLSLNTPPGEVKLTLAGCERNAGMCSLAGFTQIVNEARIPACSL 410  
 DB 363 RDKTPLSLNTPPGEVKLTLAGCERNAGMCSLAGFTQIVNEARIPACSL 412

RESULT 3  
 ABR82310  
 ID ABR82310 standard; protein; 412 AA.

XX  
 AC ABR82310;  
 XX  
 DT 06-NOV-2003 (first entry)  
 XX  
 DE Nov9X phytase polypeptide.  
 XX



Nov9X; thermotolerant; phytase; phosphate; animal food; phosphorous; enzyme.  
 Synthetic.  
 Key Location/Qualifiers  
 Misc-difference 1 /note= "encoded by GAG"  
 W0203057247-A1.  
 17-JUL-2003.  
 30-DEC-2002; 2002WO-US041785.  
 28-DEC-2001; 2001US-0344523P.  
 (SYGN ) SYNGENTA PARTICIPATIONS AG.  
 (DIVE-) DIVERSA CORP.  
 Lanahan ML, Koepf E, Kretz K;  
 WPI; 2003-598328/56.  
 N-PSDB; ACF45789.  
 Preparing thermotolerant phytase useful for preparing animal feed or human food, by expressing in microbial host cell expression cassette comprising a promoter operably linked to a nucleic acid encoding the enzyme.  
 Claim 11; Fig 1B; 79pp; English.  
 The invention relates to preparing a thermotolerant phytase (I), involving expressing in a microbial host cell an expression cassette comprising a promoter operably linked to a nucleic acid molecule encoding (I). The thermotolerant phytase has a specific activity of greater than 400 U/mg, 600 U/mg or 800 U/mg at pH 4.5 and 37plusoc. The nucleic acid molecule encodes a fusion polypeptide comprising the thermotolerant phytase and a signal sequence which is operably linked to the thermotolerant phytase. The thermotolerant phytase is glycosylated and has a simulated gastric half-life of greater than 25 minutes at a pH greater than 2.0 and less than 4.0. (I) is useful for preparing animal feed, improving nutritive value of animal feed and human food, and for preparing human food. A heat-treated animal feed mixture comprising an inorganic phosphate at below 0.45% and (I) is useful for reducing feed conversion ratios or increasing weight gain of animals fed diets with inorganic phosphate at levels below 0.45%. It is also useful for decreasing phosphate levels in excreta from an animal. An animal feed composition comprising (I) is useful for decreasing the feed conversion ratio and increasing the weight gain of an animal. The composition is useful for minimizing inorganic dietary requirements of phosphorous in an animal, for enhancing utilization of phosphorous present in feed for an animal, and for increasing organic phosphate utilization from organic phosphorus sources in feed for an animal, where the feed increases the bioavailability of inorganic phosphorus in the feed to the animal. It is also useful for decreasing phosphate levels in excreta from an animal. In all the above mentioned conditions, the feed is a poultry or swine feed. The thermotolerant phytase has a half-life of about 30 minutes in the digestive tract of the animal. The present sequence represents a thermotolerant Nov9X phytase  
 Sequence 412 AA;  
 Query Match 100.0%; Score 2153; DB 7; Length 412;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-210;  
 Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 1 QSEPELKLESVVVSRHGVRAPTKATQMQDVTDPAMTPVVKLGELTPRGELIAYLGH 60  
 3 QSEPELKLESVVVSRHGVRAPTKATQMQDVTDPAMTPVVKLGELTPRGELIAYLGH 62  
 51 YWRQLVADGLLPKCGPQSGQVAIIADVDETRTKTGEAPAAGLAPDCAITVHTQADTSS 120

Db 63 YWRQLVADGLLPKCGPQSGQVAIIADVDETRTKTGEAPAAGLAPDCAITVHTQADTSS 122  
 QY 121 PDPLFNPLKTVGCQDLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLNFPQSNCLK 180  
 Db 123 PDPLFNPLKTVGCQDLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLNFPQSNCLK 182  
 QY 181 REKQDECSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEFGGRITDS 240  
 Db 183 REKQDECSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEFGGRITDS 242  
 QY 241 HOWNTLLSLHNAQFDLIQRTPEVARSRATPLDLIKTALTTPHPKQAYGVTLPSTVLFI 300  
 Db 243 HOWNTLLSLHNAQFDLIQRTPEVARSRATPLDLIKTALTTPHPKQAYGVTLPSTVLFI 302  
 QY 301 AGHDNTNLNLGGALELNWTLPGQPDNTPPGGELVFERWRLSDNSQWTVQSVLFTQLQOM 360  
 Db 303 AGHDNTNLNLGGALELNWTLPGQPDNTPPGGELVFERWRLSDNSQWTVQSVLFTQLQOM 362  
 QY 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 410  
 Db 363 RDKTPLSLNTPPGEVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 412  
 RESULT 4  
 AAE37853  
 ID AAE37853 standard; protein; 431 AA.  
 XX AC AAE37853;  
 XX DT 06-NOV-2003 (first entry)  
 XX DE PNOV4054 phytase fusion protein.  
 XX KW Thermotolerant phytase; weight gain; animals fed diet; grain processing;  
 XX KW nutritive value; transformed plant; anabolic; chimeric; maize.  
 XX OS Chimeric - Zea mays.  
 XX OS Chimeric - Unidentified.  
 XX FH Key Location/Qualifiers  
 XX FT Peptide 1..19 /label= Signal\_peptide  
 XX FT Protein 20..431 /note= "Mature pNOV4054 phytase fusion protein"  
 XX W0203057248-A1.  
 XX 17-JUL-2003.  
 XX 30-DEC-2002; 2002WO-US041787.  
 XX 28-DEC-2001; 2001US-0344476P.  
 XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
 XX Lanahan MB, Betts S;  
 XX WPI; 2003-607980/57.  
 XX N-PSDB; AAD57148.  
 XX Preparing a thermotolerant phytase for preparing animal feed or human food by expressing in a plant cell an expression cassette comprising a promoter operably linked to a nucleic acid molecule encoding a thermotolerant phytase.  
 XX Claim 63; Page 90; 157pp; English.  
 XX The invention relates to a method for preparing a thermotolerant phytase. The method comprises expressing in a plant cell an expression cassette comprising a promoter operably linked to a nucleic acid molecule encoding a thermotolerant phytase which retains at least 40% activity after 30 minutes at 60plusoc and has a specific activity of greater than 200 U/mg at pH 4.5 and 37plusoc. The method is useful for preparing a

thermotolerant phytase for preparing animal feed or human food. The invention is useful for reducing the feed conversion ratio and increasing weight gain, improving reducing feed conversion ratios or increasing weight gain of animals fed diets with inorganic phosphate at levels below 0.4%, minimising dietary requirements of phosphorus in an animal, enhancing the utilisation of phosphorus present in animal feed, enhancing organic phosphorus utilisation from organic phosphorus sources, in animal feed, decreasing the phosphate levels in excreta from an animal, improving the processing of grain, improving the nutritive value of processed grain product or a method of processing grain, improving the nutritive value of animal feed and human food, and preparing a transgenic plant which expresses a thermotolerant phytase. The present sequence is PNOV4054 phytase fusion protein used in the exemplification of the invention

XX SQ Sequence 431 AA;

Query Match 100.0%; Score 2153; DB 6; Length 431;  
Best Local Similarity 100.0%; Pred. No. 1.5e-210;  
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QSEPELKLESVIVSRHGVRAPTKATQMDVTPDAPWTPVKLGELTPRGGELIAYLGH 60  
DB 22 QSEPELKLESVIVSRHGVRAPTKATQMDVTPDAPWTPVKLGELTPRGGELIAYLGH 81  
QY 61 YWQRLVADGLLPKCGPQSGQVAIADVDERTKTEGAFAAGLADPCAITVHTQADTSS 120  
DB 82 YWQRLVADGLLPKCGPQSGQVAIADVDERTKTEGAFAAGLADPCAITVHTQADTSS 141  
QY 121 PDPLFNPLKTVGCOLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLPFQSNCLK 180  
DB 142 PDPLFNPLKTVGCOLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLPFQSNCLK 201  
QY 181 REKQDSCSLTQALPSELKVSADCVSLTGVASLSMLTEIFLLQQAQMPGPGWGRITDS 240  
DB 202 REKQDSCSLTQALPSELKVSADCVSLTGVASLSMLTEIFLLQQAQMPGPGWGRITDS 261  
QY 241 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLDLIKTALTPHPPOKQAYGVTLPSTVLFI 300  
DB 262 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLDLIKTALTPHPPOKQAYGVTLPSTVLFI 321  
QY 301 AGHDTNLNLANLGGALELNTLPGQDNTFPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 360  
DB 322 AGHDTNLNLANLGGALELNTLPGQDNTFPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 381  
QY 361 RDKTPLSLNTPPGEVKLTLAGCERNAGQMCISLAGFTQIVNEARIPACSL 410  
DB 382 RDKTPLSLNTPPGEVKLTLAGCERNAGQMCISLAGFTQIVNEARIPACSL 431

RESULT 5  
ADC87743  
ID ADC87743 standard; protein; 436 AA.

AC ADC87743;

DT 01-JAN-2004 (first entry)

DE Modified Escherichia coli phytase.

KW Phytase; food supplement; enzyme delivery matrix; soybean meal;  
KW thermotolerance; thermostability; kernel; phytate;  
KW myo-inositol-hexaphosphate; inositol; inorganic phosphate;  
KW thermotolerant; feed value; digestion; enzyme.

OS Synthetic.

OS Escherichia coli.

PN US2003103958-A1.

PD 05-JUN-2003.

XX 24-MAY-2002; 2002US-00156660.

XX 13-AUG-1997; 97US-00910798.  
PR 01-MAR-1999; 99US-00259214.  
PR 13-APR-1999; 99US-00291931.  
PR 25-MAY-1999; 99US-00318528.  
PR 25-MAY-2000; 2000US-00580515.  
PR 24-MAY-2001; 2001US-00866379.  
XX (DIVE-) DIVERSA CORP.  
PA Short JM, Kretz K, Gray KA, Barton NR, Garrett JB, O'donoghue E;  
PI Mathur EJ;  
FI MPI; 2003-787039/74.  
XX N-PSDB; ADC87742.  
DR WPI; 2003-787039/74.  
XX N-PSDB; ADC87742.  
PT New nucleic acid encoding a polypeptide having phytase activity, useful  
PT in improving the feeding value of phytate rich ingredients or as an aid  
PT in phytate digestion.  
XX Claim 60; SEQ ID NO 2; 113pp; English.  
XX The invention discloses a new isolated or recombinant nucleic acid which  
CC encodes a polypeptide having a phytase activity. Also claimed is a  
CC nucleic acid probe, an amplification primer sequence pair, an expression  
CC cassette comprising the nucleic acid, a vector comprising the nucleic  
CC acid, a transgenic non-human animal or plant, or its seed, comprising the  
CC nucleic acid, an antisense oligonucleotide, inhibiting the translation of  
CC a phytase message in a cell, a heterodimer comprising the polypeptide and  
CC a second domain, an array comprising immobilised polypeptide or nucleic  
CC acid, a hybridoma comprising an antibody that specifically binds to the  
CC polypeptide, a food supplement for an animal, an edible enzyme delivery  
CC matrix, an edible pellet comprising a granule carrier and the  
CC polypeptide, a feed composition, a soybean meal, isolating or identifying  
CC the polypeptide, making an anti-phytase antibody, producing a recombinant  
CC polypeptide, determining whether a compound binds to the polypeptide,  
CC identifying a modulator, whole cell engineering of new or modified  
CC phenotypes by using real-time metabolic flux analysis, increasing  
CC thermotolerance or thermostability of the phytase polypeptide, increasing  
CC the resistance of the phytase polypeptide to enzymatic inactivation in a  
CC digestive system and processing of corn and sorghum kernels. The phytase  
CC activity comprises catalysis of phytate (myo-inositol-hexaphosphate) to  
CC inositol and inorganic phosphate and is thermotolerant. The nucleic acid  
CC is useful in improving the feeding value of phytate rich ingredients or  
CC as an aid in phytate digestion. The sequence presented is the modified  
CC Escherichia coli phytase.  
XX Sequence 436 AA;  
SQ Query Match 100.0%; Score 2153; DB 7; Length 436;  
Best Local Similarity 100.0%; Pred. No. 1.6e-210;  
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QSEPELKLESVIVSRHGVRAPTKATQMDVTPDAPWTPVKLGELTPRGGELIAYLGH 60  
DB 23 QSEPELKLESVIVSRHGVRAPTKATQMDVTPDAPWTPVKLGELTPRGGELIAYLGH 82  
QY 61 YWQRLVADGLLPKCGPQSGQVAIADVDERTKTEGAFAAGLADPCAITVHTQADTSS 120  
DB 83 YWQRLVADGLLPKCGPQSGQVAIADVDERTKTEGAFAAGLADPCAITVHTQADTSS 142  
QY 121 PDPLFNPLKTVGCOLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLPFQSNCLK 180  
DB 143 PDPLFNPLKTVGCOLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLPFQSNCLK 202  
QY 181 REKQDSCSLTQALPSELKVSADCVSLTGVASLSMLTEIFLLQQAQMPGPGWGRITDS 240  
DB 203 REKQDSCSLTQALPSELKVSADCVSLTGVASLSMLTEIFLLQQAQMPGPGWGRITDS 262  
QY 241 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLDLIKTALTPHPPOKQAYGVTLPSTVLFI 300  
DB 263 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLDLIKTALTPHPPOKQAYGVTLPSTVLFI 322

QY 301 AGHDTNLNLGGALELNLWTLPGQDNTPPGGELVFERWRRLSDNSQWIVSLVFQTLQOM 360  
DB 323 AGHDTNLNLGGALELNLWTLPGQDNTPPGGELVFERWRRLSDNSQWIVSLVFQTLQOM 382  
QY 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 410  
DB 383 RDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 432

RESULT 6  
AAE37854  
ID AAE37854 standard; protein; 437 AA.  
AC AAE37854;  
XX  
DT 06-NOV-2003 (first entry)  
XX  
DE pNOV4058 phytase fusion protein.  
XX  
XX Thermotolerant phytase; weight gain; animals fed diet; grain processing;  
KW nutritive value; transformed plant; anabolic; chimeric; maize.  
XX  
OS Chimeric - Zea mays.  
OS Chimeric - Unidentified.  
XX  
XX Key Location/Qualifiers  
FT Peptide 1..119  
FT /label= Signal\_peptide  
FT Protein 20..437  
FT /note= "Mature pNOV4058 phytase fusion protein"  
FT Region 432..437  
FT /note= "SEKDEL signal"  
XX  
XX WO2003057248-A1.  
XX  
XX 17-JUL-2003.  
XX  
XX 30-DEC-2002; 2002WO-US0411787.  
XX  
XX 28-DEC-2001; 2001US-0344476P.  
XX  
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX  
XX Lanahan ME, Betts S;  
XX  
XX WPI; 2003-607980/57.  
XX N-PSDB; AAD57149.  
XX  
XX Preparing a thermotolerant phytase for preparing animal feed or human  
PT food by expressing in a plant cell an expression cassette comprising a  
PT promoter operably linked to a nucleic acid molecule encoding a  
PT thermotolerant phytase.  
XX  
XX Example 1; Page 91; 157pp; English.  
XX  
XX The invention relates to a method for preparing a thermotolerant phytase.  
CC The method comprises expressing in a plant cell an expression cassette  
CC comprising a promoter operably linked to a nucleic acid molecule encoding  
CC a thermotolerant phytase which retains at least 40% activity after 30  
CC minutes at 60plusoC and has a specific activity of greater than 200 U/mg  
CC at pH 4.5 and 37plusoC. The method is useful for preparing a  
CC thermotolerant phytase for preparing animal feed or human food. The  
CC invention is useful for reducing the feed conversion ratio and increasing  
CC weight gain, improving reducing feed conversion ratios or increasing  
CC weight gain of animals fed diets with inorganic phosphate at levels below  
CC 0.45%, minimizing dietary requirements of phosphorus in an animal,  
CC enhancing the utilisation of phosphorus present in animal feed, enhancing  
CC organic phosphorus utilisation from organic phosphorus sources in animal  
CC feed, decreasing the phosphate levels in excreta from an animal,  
CC improving the processing of grain, improving the nutritive value of  
CC processed grain product or a method of processing grain, improving the  
CC nutritive value of animal feed and human food, and preparing a  
CC transformed plant which expresses a thermotolerant phytase. The present

CC sequence is pNOV4058 phytase fusion protein used in the exemplification  
CC of the invention  
XX  
SQ Sequence 437 AA;  
Query Match 100.0%; Score 2153; DB 6; Length 437;  
Best Local Similarity 100.0%; Pred. No. 1.6e-210; Mismatches 0; Gaps 0;  
Matches 410; Conservative 0; Indels 0;  
QY 1 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTPDAMPTWPKVKGELTPRGGELIAYLGH 60  
DB 22 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTPDAMPTWPKVKGELTPRGGELIAYLGH 81  
QY 61 YWRQRLVADGLLPKCGCPSQGOVATIIADVDERTRKTGEAFAAGLAPDCAITVHTQADTSS 120  
DB 82 YWRQRLVADGLLPKCGCPSQGOVATIIADVDERTRKTGEAFAAGLAPDCAITVHTQADTSS 141  
QY 121 PDPLFNPLKTVGCQLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLNPPQSNLCLK 180  
DB 142 PDPLFNPLKTVGCQLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLNPPQSNLCLK 201  
QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 240  
DB 202 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 261  
QY 241 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLLDLIKLTALTTPHPQKQAYGVTLPSTVLFI 300  
DB 262 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLLDLIKLTALTTPHPQKQAYGVTLPSTVLFI 321  
QY 301 AGHDTNLNLGGALELNLWTLPGQDNTPPGGELVFERWRRLSDNSQWIVSLVFQTLQOM 360  
DB 322 AGHDTNLNLGGALELNLWTLPGQDNTPPGGELVFERWRRLSDNSQWIVSLVFQTLQOM 381  
QY 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 410  
DB 382 RDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 431

RESULT 7  
AAE15808  
ID AAE15808 standard; protein; 430 AA.  
XX  
AC AAE15808;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Escherichia coli appA phytase mutant protein.  
XX  
XX Bacterial phytase; K12 appA phytase; protease stability; anabolic;  
KW gastrointestinal; nutritional value; feed treatment process; therapy;  
KW thermal tolerance; growth performance; alcoholic drink; biopulping;  
KW non-alcoholic drink; bioleaching; mutant; muten.  
XX  
OS Escherichia coli.  
OS Synthetic.  
XX  
XX Key Location/Qualifiers  
FH Misc-difference 68 /note= "Wild type Trp substituted with Glu"  
FT Misc-difference 84 /note= "Wild type Gln substituted with Trp"  
FT Misc-difference 84 /note= "Wild type Gln substituted with Trp"  
FT Misc-difference 95 /note= "Wild type Ala substituted with Pro"  
FT Misc-difference 97 /note= "Wild type Lys substituted with Cys"  
FT Misc-difference 168 /note= "Wild type Ser substituted with Glu"  
FT Misc-difference 180 /note= "Wild type Arg substituted with Tyr"  
FT Misc-difference 225 /note= "Wild type Asn substituted with Cys"  
FT Misc-difference 276 /note= "Wild type Tyr substituted with Asp"

XX WO200190333-A2.  
PN 29-NOV-2001.  
XX 24-MAY-2001; 2001WO-US017119.  
XX 25-MAY-2000; 2000US-00580515.  
XX (DIVE-) DIVERSA CORP.  
XX Short JM, Kretz KA, Gray KA, Barton NR, Garrett JB, O'donoghue E;  
XX WPI; 2002-083108/11.  
XX New bacterial phytase for e.g. improving the nutritional value of phytate  
XX containing foodstuffs and subsequently improving the growth performance  
XX of an organism that consumes it, or in treating animal digestive systems.  
XX Claim 59; Fig 8; 170pp; English.  
XX The patent discloses recombinant bacterial phytase from Escherichia coli  
XX K12 appA phytase. The enzyme has phytase activity and improved thermal  
XX tolerance when compared with wild-type phytase. It has improved protease  
XX stability at low pH. The recombinant phytase is useful for improving the  
XX nutritional value of phytate-containing foodstuffs and subsequently  
XX improving the growth performance of an organism that consumes it, in  
XX treating animal digestive systems, in feed treatment processes and for in  
XX vitro purposes related to research, discovery and development. They are  
XX also used for generating recombinant digestive system life forms, for  
XX the use of moulds, grains and/or plants, in biopulping and bio-bleaching  
XX where a reduction in the use of environmentally harmful chemicals that  
XX are traditionally used in the pulp and paper industry is desired and in  
XX the reduction or possible elimination of the need for mineral  
XX supplements, enzymes or therapeutic drugs for animals from the daily feed  
XX thus increasing the amount calories and nutrients present in the feed.  
XX The present sequence is E. coli appA phytase mutant protein  
XX Sequence 430 AA;  
XX  
XX Query Match 98.6%; Score 2123; DB 5; Length 430;  
XX Best Local Similarity 99.5%; Pred. No. 1.8e-207;  
XX Matches 408; Conservative 0; Mismatches 0; Indels 2; Gaps 2;  
QY 1 QSEPELKLESVIVSRHGVRAPTKATQMQVTPDANFTWPKLGLTTPRGGLIAYLGH 60  
Db 23 QSEPELKLESVIVSRHGVRAPTKATQMQVTPDANFTWPKLGLTTPRGGLIAYLGH 82  
QY 61 YWRQRLVADGILLPKCGCPQSQQVAIIADVDERTKTEGEAFAGLAPDCAITVHTQADTSS 120  
Db 83 YWRQRLVADGILLPKCGCPQSQQVAIIADVDERTKTEGEAFAGLAPDCAITVHTQADTSS 142  
QY 121 PDPLFNPLKTVGCOLDNANVTDAILERAGGSIADFTGHYQATPRELERNVFPQSNICLK 180  
Db 143 PDPLFNPLKTVGCOLDNANVTDAILE-AGGSIADFTGHYQATPRELERNVFPQSNICLK 201  
QY 181 REKQDESCLTQALPSELKVSADCVSLTGVASLSMLTEIFLLQOAGMPFGHGRITDS 240  
Db 202 REKQDESCLTQALPSELKVSADCVSLTGVASLSMLTEIFLLQOAGMPFGHGRITDS 261  
QY 241 HOWNTLISHNAQFLLORTPEVARSRAITPLDLIKTALTTPHPQKQAYGVTLTSLVFI 300  
Db 262 HOWNTLISHNAQFLLORTPEVARSRAITPLDLIKTALTTPHPQKQAYGVTLTSLVFI 320  
QY 301 AGHDTNLANLGALBELNWTLPQPDNTPPGGELVFERWRRLSDNSQWITQVSLVFTQLOOM 360  
Db 321 AGHDTNLANLGALBELNWTLPQPDNTPPGGELVFERWRRLSDNSQWITQVSLVFTQLOOM 380  
QY 361 RKPTPLSLNTPPGEVKLTLAGCEERNAQCMCSLAGTQIVNEARIPACSL 410  
Db 381 RKPTPLSLNTPPGEVKLTLAGCEERNAQCMCSLAGTQIVNEARIPACSL 430

RESULT 8  
ADA19451  
ID ADA19451 standard; protein; 430 AA.  
XX  
XX ADA19451;  
XX  
XX 20-NOV-2003 (first entry)  
XX  
XX E. coli K12 phytase mutant.  
XX  
XX Phytase; enzyme; phytate; appA gene; animal feed; inorganic phosphate;  
XX digestion enhancement; transgenic; thermal tolerance; protease stability;  
XX mutant; muten.  
XX  
XX Synthetic.  
XX Escherichia coli; strain K12.  
XX  
XX Key Location/Qualifiers  
XX Misc-difference 68 /note= "Wild-type Trp substituted by Glu"  
XX Misc-difference 72 /note= "Encoded by CGN"  
XX Misc-difference 84 /note= "Wild-type Gln substituted by Trp"  
XX Misc-difference 95 /note= "Wild-type Ala substituted by Pro"  
XX Misc-difference 97 /note= "Wild-type Lys substituted by Cys"  
XX Misc-difference 168 /note= "Wild-type Ser substituted by Glu"  
XX Misc-difference 180 /note= "Wild-type Arg substituted by Tyr"  
XX Misc-difference 226 /note= "Wild-type Asn substituted by Cys"  
XX Misc-difference 277 /note= "Wild-type Tyr substituted by Asp"  
XX  
XX US2002136754-A1.  
XX  
XX 26-SEP-2002.  
XX  
XX 24-MAY-2001; 2001US-00866379.  
XX  
XX 13-AUG-1997; 97US-00910798.  
XX 01-MAR-1999; 99US-00259214.  
XX 13-APR-1999; 99US-00291931.  
XX 25-MAY-1999; 99US-00318528.  
XX 25-MAY-2000; 2000US-00580515.  
XX  
XX (SHOR/) SHORT J M.  
XX (KRET/) KRETZ K A.  
XX (GRAY/) GRAY K A.  
XX (BART/) BARTON N R.  
XX (GAR/) GARRETT J B.  
XX (DONG/) O' DONOGHUE E.  
XX (MATH/) MATHUR E J.  
XX  
XX Short JM, Kretz KA, Gray KA, Barton NR, Garrett JB;  
XX O' Donoghue E, Mathur EJ;  
XX WPI; 2003-040002/03.  
XX  
XX Isolated Escherichia coli polynucleotide encoding a modified phytase  
XX enzyme, useful in the production of animal feed, for improving the  
XX nutritional value of phytate-containing foodstuff and for enhancing  
XX digestion in humans and animals.  
XX  
XX Claim 59; Fig 8; 62pp; English.  
XX  
XX The invention relates to an isolated Escherichia coli polynucleotide  
XX encoding a phytase enzyme appearing as ADA19450 and having amino acids  
XX modification from W68E, Q84W, A95P, K97C, S168E, R180Y, N226C or Y277D.  
XX

CC Also included the E. coli appA gene ADA19449 (or an oligonucleotide  
CC derived from it) or its mutant sequence ADA19452, expression vectors,  
CC host cells, a method of improving nutritional value of a phytate-  
CC containing foodstuff by contacting the phytate-containing foodstuff with  
CC a pure phytase enzyme mutated phytase (where the phytase enzyme catalyzes  
CC the liberation of inorganic phosphate from the phytate in the phytate-  
CC containing foodstuff), a method to produce an animal feed containing a  
CC microbial phytase (comprising culturing the plant cell, plant part or  
CC plant under conditions where the nucleotide sequence is expressed and  
CC converting the plant cells, plants or plant into a composition for animal  
CC feed), a feed composition for animals (comprising the plant seeds, plant  
CC cells, plant parts or plants in admixture with a phytate-containing  
CC foodstuff), a method to treat a human or an animal able to benefit from  
CC digestive enhancement by the activity of an exogenous phytase enzymes  
CC comprising administering to the human or animal the plant seed, plant  
CC cells, plant parts or plants of a transgenic plant which is modified to  
CC contain an expression system which expresses a nucleotide sequence  
CC encoding a phytase enzyme, a transgenic non-human organism whose genome  
CC comprising a heterologous nucleic acid sequence encoding a polypeptide  
CC having phytase activity. The phytase enzyme is useful for improving the  
CC nutritional value of phytate-containing foodstuff in the production of  
CC animal feed and for enhancing digestion in humans and animals. The  
CC invented method improves thermal tolerance and protease stability. It  
CC also improves the feeding value of phytate rich ingredients. The present  
CC sequence represents E. coli K12 mutant phytase.

XX Sequence 430 AA;

Query Match 98.6%; Score 2123; DB 6; Length 430;

Best Local Similarity 99.5%; Pred. No. 1.8e-207; Indels 2; Gaps 2;

Matches 408; Conservative 0; Mismatches 0;

1 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTDPANFTWPKLGTLPKGGELIAYLGH 60

23 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTDPANFTWPKLGTLPKGGELIAYLGH 82

61 YRQRLVADGLPKCGCPSQGVAIADVDERTKTKGEAFAGLAPDCAITVHTQADTSS 120

83 YRQRLVADGLPKCGCPSQGVAIADVDERTKTKGEAFAGLAPDCAITVHTQADTSS 142

121 PDPLENPLKTVGCQLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLFNFPQSNI 180

143 PDPLENPLKTVGCQLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLFNFPQSNI 201

181 REKQDESCLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLOQAQMPGPGGRITDS 240

202 REKQDESCLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLOQAQMPGPGGRITDS 261

241 HQWNTLLSLHNAQFD-LQRTPEVARSRAITPLDLIKTALTTPHPKQAYGVTLFTSVLFI 300

262 HQWNTLLSLHNAQFD-LQRTPEVARSRAITPLDLIKTALTTPHPKQAYGVTLFTSVLFI 320

301 AGHDTNLANLGALSLNWTLPQDPNTPPGGELVFERWRRLSDNSQIOTVSLVFQTLQOM 360

321 AGHDTNLANLGALSLNWTLPQDPNTPPGGELVFERWRRLSDNSQIOTVSLVFQTLQOM 380

361 RDKTFLSLNTPPGVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 410

381 RDKTFLSLNTPPGVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 430

RESULT 9

AAB36257

ID AAB36257 standard; protein; 432 AA.

XX AAB36257;

XX 12-SEP-2003 (revised)

DT 20-FEB-2001 (first entry)

XX Lama2/APPA plasmid translated sequence.

XX Transgenic animal; salivary protein; phytase; phosphorus; animal growth;

environmental pollution; pig.

Mus musculus.

Escherichia coli.

Chimeric.

WO200064247-A1.

02-NOV-2000.

20-APR-2000; 2000WO-CA000430.

23-APR-1999; 99US-0130508P.

(UYGU-) UNIV GUELPH.

Forberg CW, Golovan S, Phillips JP;

WPI; 2000-687245/67.

N-PSDB; AAC68294.

Transgenic non-human animal for gastrointestinal tract specific

expression of a protein, preferably phytase, comprises a nucleic acid

sequence including a heterologous transgene construct encoding the

protein.

Disclosure; Fig 5; 152pp; English.

The present invention provides transgenic animals which produce desired

proteins, in this case pigs which expresses phytase in the salivary

gland. Low phytase production levels result in phytate in the diet being

excreted and causing phosphorus contamination in water, as well as

reducing the growth of animals. The invention provides a number of

transgenes containing the E. coli APPA phytase coding sequence. (Updated

on 12-SEP-2003 to standardise OS field)

XX Sequence 432 AA;

Query Match 96.5%; Score 2077; DB 3; Length 432;

Best Local Similarity 98.0%; Pred. No. 9e-203;

Matches 402; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

1 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTDPANFTWPKLGTLPKGGELIAYLGH 60

23 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTDPANFTWPKLGTLPKGGELIAYLGH 82

61 YRQRLVADGLPKCGCPSQGVAIADVDERTKTKGEAFAGLAPDCAITVHTQADTSS 120

83 YRQRLVADGLPKCGCPSQGVAIADVDERTKTKGEAFAGLAPDCAITVHTQADTSS 142

121 PDPLENPLKTVGCQLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLFNFPQSNI 180

143 PDPLENPLKTVGCQLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLFNFPQSNI 202

181 REKQDESCLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLOQAQMPGPGGRITDS 240

203 REKQDESCLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLOQAQMPGPGGRITDS 262

241 HQWNTLLSLHNAQFD-LQRTPEVARSRAITPLDLIKTALTTPHPKQAYGVTLFTSVLFI 300

263 HQWNTLLSLHNAQFD-LQRTPEVARSRAITPLDLIKTALTTPHPKQAYGVTLFTSVLFI 322

301 AGHDTNLANLGALSLNWTLPQDPNTPPGGELVFERWRRLSDNSQIOTVSLVFQTLQOM 360

323 AGHDTNLANLGALSLNWTLPQDPNTPPGGELVFERWRRLSDNSQIOTVSLVFQTLQOM 382

361 RDKTFLSLNTPPGVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 410

383 RDKTFLSLNTPPGVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 432

RESULT 10

AAB36261

ID XX AAB36261 standard; protein; 432 AA.  
 AC AAB36261;  
 DT 12-SEP-2003 (revised)  
 DT 20-FEB-2001 (first entry)  
 DE R15/APPA plasmid translated sequence.  
 KW Transgenic animal; salivary protein; phytase; phosphorus; animal growth;  
 KW environmental pollution; pig.  
 OS Rattus sp.  
 OS Escherichia coli.  
 OS Chimeric.  
 FN WO200064247-A1.  
 PD 02-NOV-2000.  
 XX 20-APR-2000; 2000WO-CA000430.  
 XX 23-APR-1999; 99US-0130508P.  
 PA (UYGU-) UNIV GUELPH.  
 PI Forsberg CW, Golovan S, Phillips JP;  
 XX WPI; 2000-687245/67.  
 DR N-PSDB; AAC68298.  
 XX Transgenic non-human animal for gastrointestinal tract specific  
 PT expression of a protein, preferably phytase, comprises a nucleic acid  
 PT sequence including a heterologous transgene construct encoding the  
 PT protein.  
 XX Disclosure; Fig 21; 152pp; English.  
 CC The present invention provides transgenic animals which produce desired  
 CC proteins, in this case pigs which expresses phytase in the salivary  
 CC gland. Low phytase production levels result in phytate in the diet being  
 CC excreted and causing phosphorus contamination in water, as well as  
 CC reducing the growth of animals. The invention provides a number of  
 CC transgenes containing the E. coli APPA phytase coding sequence. (Updated  
 CC on 12-SEP-2003 to standardise OS field)  
 XX Sequence 432 AA;  
 SQ  
 Query Match 96.5%; Score 2077; DB 3; Length 432;  
 Best Local Similarity 98.0%; Pred. No. 9e-203;  
 Matches 402; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
 QY 1 QSEPELKLESVVVSRHGVRAFTKATQMQDVTDPAMPTWPKLGELTPRGGLIAYLGH 60  
 DB 23 QSEPELKLESVVVSRHGVRAFTKATQMQDVTDPAMPTWPKLGELTPRGGLIAYLGH 82  
 QY 61 YRQRLVADGLLPKCGCPQSGQVAILADVDETRTKTGEAFAGLAPDCAITVHTQADTSS 120  
 DB 83 YRQRLVADGLLPKCGCPQSGQVAILADVDETRTKTGEAFAGLAPDCAITVHTQADTSS 142  
 QY 121 PDPLFNPLKTGVCQLDNANVTDAILERAGGSIADFTGHYQTAFAFRELVRNFPQSNCLCK 180  
 DB 143 PDPLFNPLKTGVCQLDNANVTDAILERAGGSIADFTGHYQTAFAFRELVRNFPQSNCLCK 202  
 QY 181 REKQDESCLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLOAQMPPEGGRITDS 240  
 DB 203 REKQDESCLTQALPSELKVSADNVSLSLTVGAVSLASMLTEIFLLOAQMPPEGGRITDS 262  
 QY 241 HOWNLLSHNAQFOLLORTTEVARSATPLDLIKTALTTPHPQKQAYGVTLTSTVLF 300  
 DB 263 HOWNLLSHNAQFOLLORTTEVARSATPLDLIKTALTTPHPQKQAYGVTLTSTVLF 322  
 QY 301 AGHDTNLANLGALSNLWTLPGQPDNTPPGGLVFERWRRLSDNSQWISLVFTQLQOM 360

DB 323 AGHDTNLANLGALSNLWTLPGQPDNTPPGGLVFERWRRLSDNSQWISLVFTQLQOM 382  
 QY 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQCMCSLAGFTQIVNEARIPACSL 410  
 DB 383 RDKTPLSLNTPPGEVKLTLAGCEERNAQCMCSLAGFTQIVNEARIPACSL 432  
 RESULT 11  
 AAB36262  
 ID AAB36262 standard; protein; 432 AA.  
 AC AAB36262;  
 XX 12-SEP-2003 (revised)  
 DT 20-FEB-2001 (first entry)  
 DE SV40/APPA plasmid translated sequence.  
 KW Transgenic animal; salivary protein; phytase; phosphorus; animal growth;  
 KW environmental pollution; pig.  
 OS Simian virus 40.  
 OS Escherichia coli.  
 OS Chimeric.  
 FN WO200064247-A1.  
 PD 02-NOV-2000.  
 XX 20-APR-2000; 2000WO-CA000430.  
 XX 23-APR-1999; 99US-0130508P.  
 PA (UYGU-) UNIV GUELPH.  
 PI Forsberg CW, Golovan S, Phillips JP;  
 XX WPI; 2000-687245/67.  
 DR N-PSDB; AAC68299.  
 XX Transgenic non-human animal for gastrointestinal tract specific  
 PT expression of a protein, preferably phytase, comprises a nucleic acid  
 PT sequence including a heterologous transgene construct encoding the  
 PT protein.  
 XX Disclosure; Fig 22; 152pp; English.  
 CC The present invention provides transgenic animals which produce desired  
 CC proteins, in this case pigs which expresses phytase in the salivary  
 CC gland. Low phytase production levels result in phytate in the diet being  
 CC excreted and causing phosphorus contamination in water, as well as  
 CC reducing the growth of animals. The invention provides a number of  
 CC transgenes containing the E. coli APPA phytase coding sequence. (Updated  
 CC on 12-SEP-2003 to standardise OS field)  
 XX Sequence 432 AA;  
 SQ  
 Query Match 96.5%; Score 2077; DB 3; Length 432;  
 Best Local Similarity 98.0%; Pred. No. 9e-203;  
 Matches 402; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
 QY 1 QSEPELKLESVVVSRHGVRAFTKATQMQDVTDPAMPTWPKLGELTPRGGLIAYLGH 60  
 DB 23 QSEPELKLESVVVSRHGVRAFTKATQMQDVTDPAMPTWPKLGELTPRGGLIAYLGH 82  
 QY 61 YRQRLVADGLLPKCGCPQSGQVAILADVDETRTKTGEAFAGLAPDCAITVHTQADTSS 120  
 DB 83 YRQRLVADGLLPKCGCPQSGQVAILADVDETRTKTGEAFAGLAPDCAITVHTQADTSS 142  
 QY 121 PDPLFNPLKTGVCQLDNANVTDAILERAGGSIADFTGHYQTAFAFRELVRNFPQSNCLCK 180  
 DB 143 PDPLFNPLKTGVCQLDNANVTDAILERAGGSIADFTGHYQTAFAFRELVRNFPQSNCLCK 202

QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMPEPGWGRITDS 240  
DB 203 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMPEPGWGRITDS 262  
QY 241 HQWNTLLSLHNAQFDLLQRTPEVARSRAATPLDLIKTALTTEHPPOKQAYGVTLPSTVLF 300  
DB 263 HQWNTLLSLHNAQFVLLQRTPEVARSRAATPLDLIKTALTTEHPPOKQAYGVTLPSTVLF 322  
QY 301 AGHDTNLNLGGALELNWTLPGQPDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 360  
DB 323 AGHDTNLNLGGALELNWTLPGQPDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 382  
QY 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 410  
DB 383 RDKTPLSLNTPPGEVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 432

RESULT 12  
AAB36259  
ID AAB36259 standard; protein; 432 AA.  
XX  
AC AAB36259;  
XX  
DT 12-SEP-2003 (revised)  
DT 20-FEB-2001 (first entry)  
XX  
DE R15/APPA plasmid translated sequence.  
XX  
XX Transgenic animal; salivary protein; phytase; phosphorus; animal growth;  
XX environmental pollution; pig.  
XX  
OS Rattus sp.  
OS Escherichia coli.  
OS Chimeric.  
XX  
FN WO200064247-A1.  
XX  
PD 02-NOV-2000.  
XX  
PF 20-APR-2000; 2000WO-CA000430.  
XX  
PR 23-APR-1999; 99US-0130508P.  
XX  
PA (UYGU-) UNIV GUELPH.  
XX  
PI Forsberg CW, Golovan S, Phillips JP;  
XX  
XX WPI; 2000-687245/67.  
DR N-PSDB; AAC68296.  
XX  
XX Transgenic non-human animal for gastrointestinal tract specific  
PT expression of a protein, preferably phytase, comprises a nucleic acid  
PT sequence including a heterologous transgene construct encoding the  
PT protein.  
XX  
XX Disclosure; Fig 19; 152pp; English.  
XX  
XX The present invention provides transgenic animals which produce desired  
CC proteins, in this case pigs which expresses phytase in the salivary  
CC gland. Low phytase production levels result in phytate in the diet being  
CC excreted and causing phosphorus contamination in water, as well as  
CC reducing the growth of animals. The invention provides a number of  
CC transgenes containing the E. coli APPA phytase coding sequence. (Updated  
CC on 12-SEP-2003 to standardise OS field)  
XX  
SQ Sequence 432 AA;  
XX  
Query Match 96.5%; Score 2077; DB 3; Length 432;  
Best Local Similarity 98.0%; Pred. No. 9e-203;  
Matches 402; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
QY 1 QSEPELKESVIVSRHGVRATPKATQLMQDVTDPANPTFWKLGWLTFRGGELIAYLGH 60

DB 23 QSEPELKESVIVSRHGVRATPKATQLMQDVTDPANPTFWKLGWLTFRGGELIAYLGH 82  
QY 61 YWRQRLVADGLLPKCCGPOSQVAVIADVDERTRKTGEFAAGLAPDCAITVHTQADTSS 120  
DB 83 YQQRVLVADGLLAKKGCPOSQVAVIADVDERTRKTGEFAAGLAPDCAITVHTQADTSS 142  
QY 121 PDLFENPLKTVGCQOLDNANVTDAILLERAGGSTADFTGHYQTAFARELERVLPNQSNLCLK 180  
DB 143 PDLFENPLKTVGCQOLDNANVTDAILLERAGGSTADFTGHYQTAFARELERVLPNQSNLCLK 202  
QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMPEPGWGRITDS 240  
DB 203 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMPEPGWGRITDS 262  
QY 241 HQWNTLLSLHNAQFDLLQRTPEVARSRAATPLDLIKTALTTEHPPOKQAYGVTLPSTVLF 300  
DB 263 HQWNTLLSLHNAQFVLLQRTPEVARSRAATPLDLIKTALTTEHPPOKQAYGVTLPSTVLF 322  
QY 301 AGHDTNLNLGGALELNWTLPGQPDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 360  
DB 323 AGHDTNLNLGGALELNWTLPGQPDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 382  
QY 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 410  
DB 383 RDKTPLSLNTPPGEVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 432

RESULT 13  
AAB36258  
ID AAB36258 standard; protein; 432 AA.  
XX  
AC AAB36258;  
XX  
DT 12-SEP-2003 (revised)  
DT 20-FEB-2001 (first entry)  
XX  
DE R15/APPA plasmid translated sequence.  
XX  
XX Transgenic animal; salivary protein; phytase; phosphorus; animal growth;  
XX environmental pollution; pig.  
XX  
OS Rattus sp.  
OS Escherichia coli.  
OS Chimeric.  
XX  
FN WO200064247-A1.  
XX  
PD 02-NOV-2000.  
XX  
PF 20-APR-2000; 2000WO-CA000430.  
XX  
PR 23-APR-1999; 99US-0130508P.  
XX  
PA (UYGU-) UNIV GUELPH.  
XX  
PI Forsberg CW, Golovan S, Phillips JP;  
XX  
XX WPI; 2000-687245/67.  
DR N-PSDB; AAC68295.  
XX  
XX Transgenic non-human animal for gastrointestinal tract specific  
PT expression of a protein, preferably phytase, comprises a nucleic acid  
PT sequence including a heterologous transgene construct encoding the  
PT protein.  
XX  
XX Disclosure; Fig 18; 152pp; English.  
XX  
XX The present invention provides transgenic animals which produce desired  
CC proteins, in this case pigs which expresses phytase in the salivary  
CC gland. Low phytase production levels result in phytate in the diet being  
CC excreted and causing phosphorus contamination in water, as well as  
CC reducing the growth of animals. The invention provides a number of  
CC



```
CC transgenes containing the E. coli APPA phytase coding sequence. (Updated
CC on 12-SEP-2003 to standardise OS field)
XX
SQ Sequence 432 AA;

  Query Match      96.5%; Score 2077; DB 3; Length 432;
  Best Local Similarity 98.0%; Pred. No. 9e-203;
  Matches 402; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

  1 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTPDAMPTWPKLGELTPRGGLIAYLGH 60
  23 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTPDAMPTWPKLGELTPRGGLIAYLGH 82
  61 YWRQRLVADGLLPKCGPQSGQVAIIADVDERTRKTGEAFAAGLAPDCAITVHTQADTSS 120
  83 YQRLVADGLLPKCGPQSGQVAIIADVDERTRKTGEAFAAGLAPDCAITVHTQADTSS 142
  121 PDPLFNLKTVGCQLDNANVTDAILERAGGSIADFTGHYQTAFARELERVLPNPPQSNLCLK 180
  143 PDPLFNLKTVGCQLDNANVTDAILERAGGSIADFTGHYQTAFARELERVLPNPPQSNLCLK 202
  181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMGPEPGWGRITDS 240
  203 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMGPEPGWGRITDS 262
  241 HQWNTLLSLHNAQFVLLQRTPEVARSRATPLLDLIKTAITLTPHPPKQAYGVTLPTSVLFI 300
  263 HQWNTLLSLHNAQFVLLQRTPEVARSRATPLLDLIKTAITLTPHPPKQAYGVTLPTSVLFI 322
  301 AGHDTNLANLGGALELNWTLPGQPDNTPPGGELVFERWRLSDNSQWISLVFQTLQOM 360
  323 AGHDTNLANLGGALELNWTLPGQPDNTPPGGELVFERWRLSDNSQWISLVFQTLQOM 382
  361 RDKTPLSLNTPPGVEVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 410
  383 RDKTPLSLNTPPGVEVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 432

RESULT 14
AAB36263
ID AAB36263 standard; protein; 432 AA.
XX
AC AAB36263;
XX
XX Mus musculus.
OS Escherichia coli.
OS Chimeric.
XX
PN WO200064247-A1.
XX
PD 02-NOV-2000.
XX
PF 20-APR-2000; 2000WO-CA000430.
XX
PR 23-APR-1999; 99US-0130508P.
XX
PA (UYGU-) UNIV GUELPH.
XX
PI Forsberg CW, Golovan S, Phillips JP;
XX
XX WPI; 2000-687245/67.
XX
DR N-PSDB; AAC68300.
XX
PT Transgenic non-human animal for gastrointestinal tract specific
PT expression of a protein, preferably phytase, comprises a nucleic acid
```

```
PT sequence including a heterologous transgene construct encoding the
PT protein.
XX
XX Disclosure; Fig 23; 152pp; English.
XX
CC The present invention provides transgenic animals which produce desired
CC proteins, in this case pigs which expresses phytase in the salivary
CC gland. Low phytase production levels result in phytate in the diet being
CC excreted and causing phosphorus contamination in water, as well as
CC reducing the growth of animals. The invention provides a number of
CC transgenes containing the E. coli APPA phytase coding sequence. (Updated
CC on 12-SEP-2003 to standardise OS field)
XX
SQ Sequence 432 AA;

  Query Match      96.5%; Score 2077; DB 3; Length 432;
  Best Local Similarity 98.0%; Pred. No. 9e-203;
  Matches 402; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

  1 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTPDAMPTWPKLGELTPRGGLIAYLGH 60
  23 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTPDAMPTWPKLGELTPRGGLIAYLGH 82
  61 YWRQRLVADGLLPKCGPQSGQVAIIADVDERTRKTGEAFAAGLAPDCAITVHTQADTSS 120
  83 YQRLVADGLLPKCGPQSGQVAIIADVDERTRKTGEAFAAGLAPDCAITVHTQADTSS 142
  121 PDPLFNLKTVGCQLDNANVTDAILERAGGSIADFTGHYQTAFARELERVLPNPPQSNLCLK 180
  143 PDPLFNLKTVGCQLDNANVTDAILERAGGSIADFTGHYQTAFARELERVLPNPPQSNLCLK 202
  181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMGPEPGWGRITDS 240
  203 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMGPEPGWGRITDS 262
  241 HQWNTLLSLHNAQFVLLQRTPEVARSRATPLLDLIKTAITLTPHPPKQAYGVTLPTSVLFI 300
  263 HQWNTLLSLHNAQFVLLQRTPEVARSRATPLLDLIKTAITLTPHPPKQAYGVTLPTSVLFI 322
  301 AGHDTNLANLGGALELNWTLPGQPDNTPPGGELVFERWRLSDNSQWISLVFQTLQOM 360
  323 AGHDTNLANLGGALELNWTLPGQPDNTPPGGELVFERWRLSDNSQWISLVFQTLQOM 382
  361 RDKTPLSLNTPPGVEVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 410
  383 RDKTPLSLNTPPGVEVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 432

RESULT 15
AAB36260
ID AAB36260 standard; protein; 432 AA.
XX
AC AAB36260;
XX
XX Transgenic animal; salivary protein; phytase; phosphorus; animal growth;
XX environmental pollution; pig.
DT 12-SEP-2003 (revised)
DT 20-FEB-2001 (first entry)
XX
DE R15/APPA plasmid translated sequence.
XX
KW Transgenic animal; salivary protein; phytase; phosphorus; animal growth;
KW environmental pollution; pig.
XX
XX Rattus sp.
OS Escherichia coli.
OS Chimeric.
XX
XX WO200064247-A1.
XX
XX 02-NOV-2000.
XX
XX 20-APR-2000; 2000WO-CA000430.
XX
XX 23-APR-1999; 99US-0130508P.
```

XX (UYGU-) UNIV GUELPH.  
XX Foreberg CW, Golovan S, Phillips JP;  
XX WPI; 2000-687245/67.  
XX N-PSDB; AAC68297.  
XX Transgenic non-human animal for gastrointestinal tract specific  
XX expression of a protein, preferably phytase, comprises a nucleic acid  
XX sequence including a heterologous transgene construct encoding the  
XX protein.  
XX Disclosure; Fig 20; 152pp; English.  
XX The present invention provides transgenic animals which produce desired  
XX proteins, in this case pigs which expresses phytase in the salivary  
XX gland. low phytase production levels result in phytate in the diet being  
XX excreted and causing phosphorus contamination in water, as well as  
XX reducing the growth of animals. The invention provides a number of  
XX transgenes containing the E. coli APPA phytase coding sequence. (Updated  
XX on 12-SEP-2003 to standardise OS field)  
XX Sequence 432 AA;  
XX  
Query Match 96.5%; Score 2077; DB 3; Length 432;  
Best Local Similarity 98.0%; Pred. No. 9e-203; 8; Indels 0; Gaps 0;  
Matches 402; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
QY 1 QSEPELKLESVIVSRHGVRAPTKATQMQDVTTPDAWTPVVKLGELTPRGGELIAYLGH 60  
Db 23 QSEPELKLESVIVSRHGVRAPTKATQMQDVTTPDAWTPVVKLGELIAYLGH 82  
QY 61 YWQRLVADGLLPKCCPOSGQVAIIADVDERTKGTGEAFAAGLAPDCAITVHTQADTSS 120  
Db 83 YQRLVADGLLAKGCPQSGQVAIIADVDERTKGTGEAFAAGLAPDCAITVHTQADTSS 142  
QY 121 PDPLFNPLKTVGCQOLDNANVTDAILRAGGSIAFTGHYQTAFRELERVLPQSNCLIK 180  
Db 143 PDPLFNPLKTVGCQOLDNANVTDAILRAGGSIAFTGHYQTAFRELERVLPQSNCLIK 202  
QY 181 REKQDESCSLTOALPSBLKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 240  
Db 203 REKQDESCSLTOALPSBLKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 262  
QY 241 HQWNTLLSHNAQFDLQRTPEVARSRATELLDLIKTALTTPHPQKQAYGVTLPTSVLFI 300  
Db 263 HQWNTLLSHNAQFDLQRTPEVARSRATELLDLIKTALTTPHPQKQAYGVTLPTSVLFI 322  
QY 301 AGHDTNLNLGGALELNTLPGQPDNTTPGGELVFERWRRLSDNSQWIQVSLVFQTLQOM 360  
Db 323 AGHDTNLNLGGALELNTLPGQPDNTTPGGELVFERWRRLSDNSQWIQVSLVFQTLQOM 382  
QY 361 RDKTPLSLNTPPGVEVKLTLAGCERNAGQCSLAGFTQIVNEARIACSL 410  
Db 383 RDKTPLSLNTPPGVEVKLTLAGCERNAGQCSLAGFTQIVNEARIACSL 432